

Collisions between science and media

Premature release of scientific results to the public is against the interests of scientists, journalists and the public. The use of embargoed press releases is beneficial, and helps the quality of coverage of science in the media.

THEY did it with 'cold fusion' and they did it with 'life on Mars': in both cases, the mass media jumped the gun on scientific journals. And did it matter? Yes, because in both cases, good media coverage required analysis of what was being claimed, necessitating informed comments from other scientists. Furthermore, in both cases there was good cause for scepticism, even beyond the general principle that grand claims require exceptionally resilient evidence. In cold fusion, the predictions of quantum mechanics were apparently being contradicted. With Martian meteoritic hydrocarbons, the parallels with past discounted evidence of early life on Earth are obvious: the context of the evidence — ancient rocks subject to diverse thermal and chemical influences — can all too easily give rise to misleadingly suggestive signatures. But, in both cases, the lack of the relevant published material made prompt informed comment by outsiders difficult.

The benefits of peer review by journals as a means of giving journalists some confidence in new work are self-evident. Premature release to the media thus denies them that confidence as well as the ability to obtain informed reactions. But conflicting pressures sometimes override the interests embodied in the formal publication process. Scientists wishing to attract attention to themselves in order to influence funding agencies, and agencies wishing to influence governments through public pressure, may on occasion ride roughshod over the proprieties of scientific publication to achieve their ends. And in another science-and-media saga, that of bovine spongiform encephalopathy (BSE), governments have not hesitated to release information prematurely when it suited them. But such actions can backfire when apparently strong conclusions tumble before close scrutiny in the peer-review process.

Such stories are of international interest and carry significant news value: interested journalists are correspondingly numerous and determined. In the case of BSE, for example, some scientists have had virtually to disconnect themselves from the press to pursue their research effectively. All the more reason, therefore, for them, their institutions and journals to release information in a controlled way. That provides equal access to the press and, if the news is released a day or two ahead of publication, journalists can give their coverage the depth that it deserves — both they and the scientists benefit from the quality that is thus attainable.

The quoted examples are extreme, but their moral applies to less sensational science as well: a press release with an embargo — a deadline before which nothing may be published — usually serves the interests of all involved. And, of course, it allows journalists, too, to enhance their public profile.

For all those reasons, and like several other journals, *Nature* has refused to publish papers prematurely released to the press, and puts out its weekly release of summarized content to journalists, embargoed until the evening before the Thursday of publication. And the embargoes are taken seriously — those who break them have been removed from the circulation list.

Hindering communication between scientists is another matter, which is why different guidelines apply to work that has been discussed at a scientific conference and picked up by the media in the process. Conferences, of course, do not constitute prior publication — their contents are not immediately available to the wider community and the peer-review process of conferences is at best

less thorough than for journals. (A substantial fraction of papers presented at conferences never appear in journals.)

Nature's guidelines for potential authors at such occasions are clear-cut in principle: talk to other researchers as much as you wish, but do not encourage or risk premature publication by discussion with the press, beyond your formal presentation. That advice may jar with those (including most researchers and all journalists) who see the freedom of information as a good thing. But it embodies a longer-term view: that publication in a peer-reviewed journal is the appropriate culmination of any piece of original research, and an essential prerequisite for public discussion. □

Don't ban sceptics

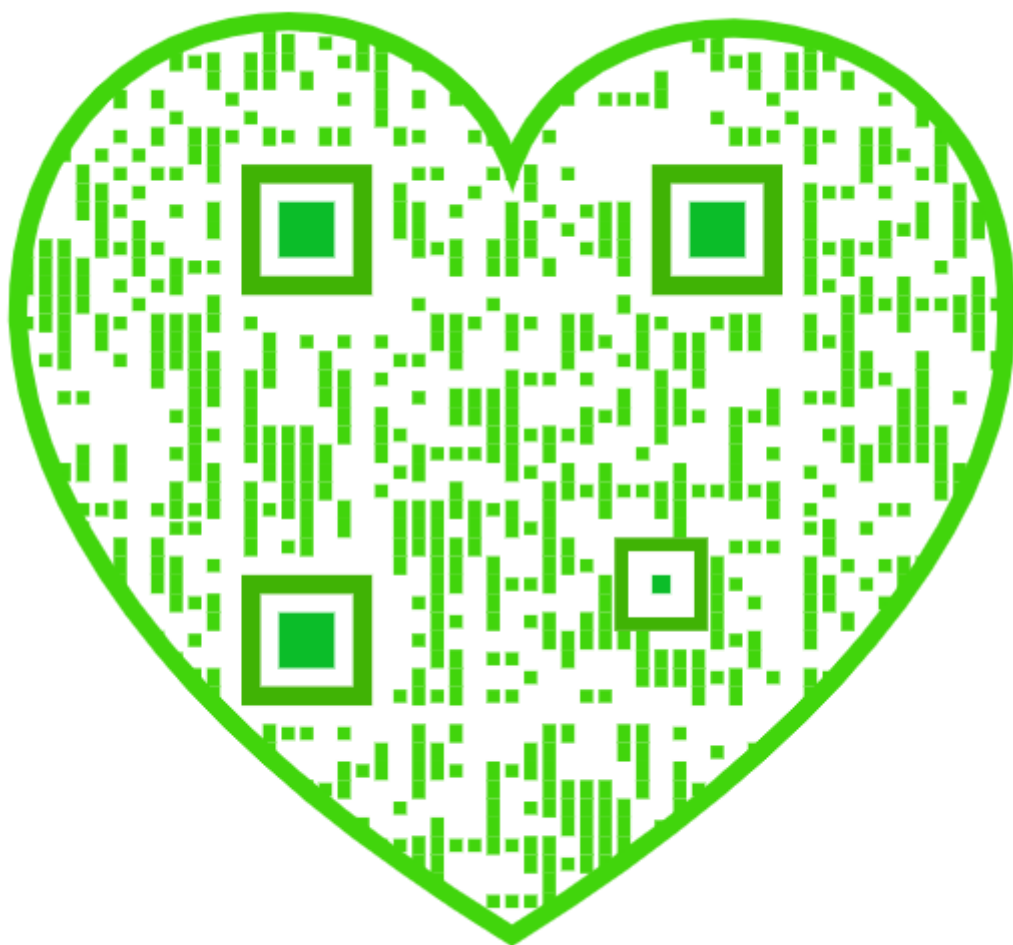
Climatological lobbyists have given scepticism a bad name, and more credible criticism is at risk of suppression.

DURING a series of lively hearings last autumn before a subcommittee of the US House of Representatives, members of the Republican majority sought to find hard evidence that the peer-review process for environmental sciences is being corrupted by institutional bias. Allegations were made to the effect that scientists who held views at variance with those underpinning the policies of the Clinton administration were being unfairly treated. It was said that unnamed government officials had flatly stated that research with certain likely outcomes would not be funded and — in later correspondence — that this journal had treated certain submissions unfairly.

But names were not named and the allegations appear to have fizzled out. Dana Rohrabacher, chair of the energy and environment subcommittee of the House Science Committee, who instigated the hearings, had every political incentive to follow up the accusations, but was unable to do so.

To that extent, George Brown (Democrat, California), the senior minority member on the Science Committee, is right to say that the hearings utterly failed to substantiate claims that biased peer review is corrupting the environmental sciences (see page 749). Brown is worried that the hearings, which called as witnesses established 'sceptics' about ozone depletion, global warming and dioxins, lent them too much credibility. He contends that by calling the 'sceptics' as witnesses next to scientists holding mainstream views, the hearings create a perception that no reliable process exists to resolve scientific differences of opinion. He fears that peer review will thus be discredited, and policy-making undermined.

It is odd for a liberal such as Brown to argue against the presentation of diverse views. The Science Committee lacks real power, but still serves as a valuable forum for science issues in the Congress. The voices of the sceptics may grow tiresome, but the mainstream is in trouble if it cannot win a public debate with them. If Brown resumes the chairmanship of the Science Committee, as he may do after next week's elections, he should make sure that the committee continues to hear from credible witnesses who are prepared to dispute the prevailing wisdom on environmental issues: there are plenty of them around. □



NASA panel gives top priority to international satellite projects

Washington. The US National Aeronautics and Space Administration (NASA) should give highest priority in its lean mission operations budget to increasing participation in several non-US astrophysics missions, says an external scientific review panel.

The panel also recommends that NASA should scale down funding for two US missions that are past their design lifetimes — the Extreme Ultraviolet Explorer (EUVE) and Compton Gamma Ray Observatory (GRO).

NASA asked the 'senior review' team, led by Andrea Dupree of the Harvard-Smithsonian Astrophysics Observatory, to rank the science merit of eight astrophysics missions whose funding was set to expire in the next few years. The space agency does not have enough money to keep all its astronomy satellites operating indefinitely, and faces the prospect of shutting down some spacecraft that are still returning useful data (see *Nature* **382**, 193; 1996).

The review team gave its top ranking to US participation in Europe's Infrared Space Observatory (ISO), which it called the "major infrared mission of this decade". The panel said "the scientific return [to US scientists] can be greatly leveraged with further modest investments" in data analysis for US guest observers, science operations, and data archiving. US astronomers have already been awarded 34 per cent of the available time on ISO.

Guenter Riegler, chief scientist for NASA's research programmes, says the agency will follow this and all other senior review recommendations, and has already notified project heads of what funding they can expect. ISO funding for 1997 to 1999 will be increased over the levels that had been planned, and will be extended (at a reduced level) to 2000.

The review rated US participation in two X-ray missions — Japan's ASCA (Advanced Satellite for Cosmology and Astrophysics) and Germany's ROSAT (Roentgen Satellite) — as the second and third most scientifically valuable areas of spending, respectively. Both will receive extended NASA funding. ASCA funding will be extended from 1997 to 2000, and ROSAT's from 1997 to 1999.

A US-only mission, the Rossi X-Ray Timing Explorer (XTE), will also be extended by one year, to 2000. But two other US missions, EUVE and GRO, received low science rankings, and their funding will suffer as a result. The panel found that extending EUVE for a second time (its lifetime has been increased once) is "not as compelling" as funding other missions in the competition, in part because its brightest targets already have been measured, and the remaining ones are harder to observe.

The GRO too is "entering an era of diminishing science returns", with the notable exception of an instrument to detect

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Flying high: US participation in Europe's ISO project won the top 'science merit' ranking.

gamma-ray bursts. Although NASA's overall funding for GRO will drop to less than 20 per cent of its original level in 1998, the burst experiment will receive the lion's share of remaining funds.

The low ranking for EUVE came as "a disappointment", says Riegler, because NASA has high-profile plans to use it as an experiment in "outsourcing" satellite operations to cut costs. The space agency recently signed an agreement with the University of California at Berkeley's Center for Extreme Ultraviolet Astrophysics (CEA) whereby the university will take over full operation of EUVE next February. Michael Gunter, EUVE project manager at Berkeley, says the satellite will be run by seven or eight people, compared to the 17 or so who now operate it for NASA. And the annual operations costs will be cut to about \$1.3 million.

NASA will still fund EUVE operations up to September 1998, but will back out after that. Roger Malina, director of the CEA, says there are no "realistic options" for maintaining the satellite at its current level of scientific productivity after NASA withdraws. Gunter says the project is exploring very reduced modes of operation, with students taking over some functions.

The senior review team acknowledged the unpleasantness of its duty, and called the prospect of shutting off working satellites "sobering". "I'd like to keep them all," says Dupree. She adds that, while NASA's participation in European or Japanese satellite projects apparently gives American scientists the highest "science return per dollar", there is some risk to this strategy. It removes US scientists and engineers from the work of building spacecraft, leading to a loss of jobs and expertise.

Tony Reichhardt

TIGER tracks down tropical carbon fix

London. A £20-million (US\$32-million) UK government environment research programme that claims to have "overturned received wisdom" to show that tropical forests can be a net absorber of carbon has come to the end of its five-year life.

The Terrestrial Initiative in Global Environmental Research (TIGER) involved 300 researchers from a range of disciplines at 60 universities and research institutes. The initiative, funded by the Natural Environment Research Council, concentrated on research into the land carbon cycle, trace greenhouse gases, energy and water budgets, and the impacts on ecosystems of environmental change.

TIGER activities took place at sites in the United Kingdom, Brazil, Cameroon, Canada, Niger and Switzerland.

Some of TIGER's most visible work related to carbon and tropical forests. The world's land areas are thought to absorb about 1.3 billion tonnes of carbon annually. But the eventual destination of this carbon is unknown, and it is thus referred to as the 'missing sink'.

TIGER researchers set out to study the carbon balance of tropical forests in Amazonia and Cameroon. Conventional knowledge suggests that forests are in a steady state — that their uptake of carbon from the atmosphere is balanced by losses from respiration and leaching.

Using a 50-foot tower built in the Amazonian rainforest and mounted with special monitoring equipment, TIGER researchers discovered that the Brazilian Amazonia is not in a steady state, but it absorbs carbon in quantities comparable in size to the 'missing sink'. □

Delays in alternative tests defer animal ban

Utrecht. The introduction of a new European Union law banning the avoidable use of animals in toxicity tests for cosmetics is likely to be delayed until 2000 because no suitable alternative tests have been successfully developed.

Member states of the union had agreed to a European Commission directive that would have banned animal tests where alternative tests exist from 1998. But animal welfare groups and industrialists attending the second World Congress on Alternatives and Animal Use in the Life Sciences in Utrecht, the Netherlands, last week reluctantly agreed that animals will continue to be used for several years to come.

The congress heard that a significant step forward had been achieved, however, with the completion of the first attempt to produce international standards for validating alternative tests. A workshop run by the Organization for Economic Cooperation and Development (OECD) has looked at the harmonization of the acceptance criteria for alternative toxicological testing. Its report was presented at the congress for discussion and approval.

But the move probably comes too late to stop the commission from invoking a clause

in its directive that provides for a two-year delay in introducing the ban if no alternative tests have been validated by January 1997. Internal commission documents indicate that it will recommend such a delay before the end of this year.

Alternative tests have proved difficult to

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Practice without pain: an artificial rat which has been developed for training in microsurgical techniques.

develop for two main reasons. First, scientists are hampered by the lack of strong data on animal tests currently used, because industry is reluctant, for commercial reasons, to make its files available to independent test developers. Industry is also reluctant to provide outsiders with the

chemicals they need to validate new tests. Second, alternative tests have to prove equal to, or better than, animal tests, even though most animal tests have never been shown to be the best models of human toxicity. Using an endpoint achieved in a complex system like an animal, which may well be less than ideal, as a reference makes validation of alternative tests more difficult than it need be, argue campaigners.

In addition, the political climate has not supported, until recently, a concerted effort for the development of alternative methods. In the past few years, however, several institutions have been founded in different countries to support such efforts and validate new tests. The European Commission established its own centre, the European Centre for the Validation of Alternative Methods at Ispra, Italy, in 1991.

Steve McIvor, head of campaigns for the international cosmetics retailer Body Shop, whose products are not tested on animals, says that replacements are not in sight for the most commonly used toxicity tests — the guinea-pig skin-sensitization test, rabbit-eye irritancy test (the Draize test), and guinea-pig and rabbit skin-irritancy tests.

Alternative-test developers have reached a dead end, he says. Along with other animal welfare campaigners, he believes regulators must accept that new methods should be developed from scratch, "without reference to existing animal tests which are both complex and of uncertain accuracy". Alternative-test developers are now also trying to acquire more human data from poison clinics, a source of information that has been insufficiently tapped, he says.

The Body Shop and other organizations are campaigning for a moratorium on the testing and introduction of new cosmetic ingredients until alternative tests are found. Industry is unlikely to agree. It is, however, just as interested as animal welfare campaigners in encouraging the development of alternative testing methods, which would be cheaper and easier to automate.

The number of animals used in cosmetics testing is not known because of difficulties in gathering reliable statistics. In Europe, figures have been published only for France and the United Kingdom. These are 20,000 (1993) and 3,500 (1994) respectively.

The OECD working group report is expected to be adopted into the organization's general toxicity testing guidelines. It may help to spread the high costs of running validation studies — large-scale blinded trials to prove the merit of alternative tests — by encouraging industry and regulatory agencies worldwide to cooperate with each other.

Alison Abbott

Energy labs link up genome efforts

Washington. The US Department of Energy (DoE) is to consolidate its \$35-million-a-year gene-sequencing effort under a single organization, the Joint Genome Institute, which it says will sequence up to 40 per cent of the 3 billion base pairs in the human genome by 2005.

Officials say the move is intended to be much more than a cosmetic change to DoE's gene-sequencing effort. It is seen as a much-needed attempt by the department to regain the initiative in the US Human Genome Project, which originated in DoE laboratories in 1986 but is now 70 per cent funded by the National Institutes of Health (NIH).

The new institute will control sequencing work at the Lawrence Berkeley and Lawrence Livermore National Laboratories in California, and at the Los Alamos National Laboratory in New Mexico. Its director, Elbert Branscomb, formerly a senior scientist at Livermore, does not deny that the DoE has been falling behind the NIH and the private sector in sequencing the genome. "We became persuaded that, if we didn't do something, that would rapidly become the case," he says.

The institute will account for about half of the DoE's total effort under the Human Genome Project. Branscomb says that it will change the way in which the

laboratories do sequencing "to a fairly dramatic extent", with all the work "being managed and funded as a single effort". He promises that it "will not be 'business as usual' in any sense at all".

He says the institute will work with overseas agencies and the NIH's National Center for Human Genome Research to compile the complete human genome sequence of 3 billion base pairs, and make it available to all researchers. Most privately-backed efforts to complete the sequence are going on "behind a proprietary wall", reinforcing the need for the public-sector effort.

The institute will aim to produce a share of the human genome proportional to its budget, which, Branscomb says, will mean between 20 and 30 per cent of the full sequence. The rest will come from NIH and non-US contributors. High overhead costs of working in the DoE laboratories will be one of the institute's main challenges.

The DoE often struggles to win recognition from the public and Congress for its contributions to health research. The laboratories jealously guard their own territories. But the directors of the three labs involved in the institute support its creation enthusiastically, Branscomb says, as a way to increase their gene-sequencing productivity.

Colin Macilwain

Congress hearings used as 'scientific court'

Washington. George Brown, the senior Democrat on the Science Committee in the US House of Representatives, has launched a blistering attack on his Republican opponents for their conduct of congressional hearings at which well-known 'sceptics' testified on the integrity of the science associated with key environmental issues.

The three hearings, entitled "Scientific Integrity and Public Trust", dealt respectively with ozone depletion, global warming and the danger of dioxins. They were called by Dana Rohrabacher (Republican, California), chair of the committee's energy and environment subcommittee, to establish if public science agencies were biased on these issues (see *Nature* 387, 329; 1995).

In a 64-page report published last week, Brown argues that the hearings not only failed to find evidence for such bias but also "constituted an unprecedented assault on the peer-review system and the scientific process itself". Brown says that the hearings "repudiated peer review" by exhibiting the view "that scientific truth is more likely to be found at the fringes of science rather than at the centre". He accuses Rohrabacher of seeking to set up "a scientific court where the subcommittee would determine scientific truth through testimony and questions".

Rohrabacher hit back quickly at Brown's allegations. "It is the very fairness of these hearings that infuriated the left-wing ideologues" among Brown's staff, he said in a statement. "After 40 years of autocratic, one-party rule, they are horrified that diverse points of view within the scientific community have been given a fair hearing, with panels of witnesses equally divided."

The hearings, which took place late last year, gave a platform to well-known 'sceptics' such as Sallie Baliunas of Harvard University and Fred Singer, an emeritus professor of the University of Virginia, who testified against the prevailing scientific con-

sensus on ozone depletion. Patrick Michaels, also of the University of Virginia, criticized prominent scientists holding mainstream views on global warming.

Brown argues that neither Baliunas nor Singer has published any recent peer-reviewed research on ozone depletion. His report also says that the 'sceptics' were unable to substantiate their most serious charge, namely that science agencies and the scientific process were systematically excluding their views from consideration.



Brown: 'sceptics' gave no evidence of bias.

Baliunas testified at the ozone hearing that she had been told "by officers of federal funding agencies" not to apply for funds "on certain questions" because "answering these questions would undermine the possibility of getting new funds" and "might deter policy-makers from 'doing the right thing'".

The report says that Baliunas was asked to name the officials in question and to provide information to back up this allegation, but declined to do so because of a fear of legal liability. In correspondence with the committee, however, Baliunas alleged that some of her work had been rejected for publication in *Nature* on political grounds. In response to an inquiry from Brown, John Maddox, then editor of *Nature*, said that the decision to reject Baliunas's paper was "made on purely technical grounds, which the authors had never chosen to rebut".

Brown says that the hearings "failed to produce credible substantiation for any of these claims of scientific misconduct" but instead "showed science being conducted in an objective and apolitical manner, con-

sistent with the traditional norms of scientific integrity". He calls on the scientific community to do more to counter publicly the positions of the sceptics.

But Rohrabacher's staff insist that the hearings were successful in highlighting the question of how agencies such as the Environmental Protection Agency conduct their peer review of science. According to one Republican staff member, government-funded scientists remain scared to testify against mainstream positions on such questions as the man-made nature of global warming.

Asked to comment on the report, Singer said that he had repeatedly published peer-reviewed research on ozone, and described the report as "a campaign to smear people". Michaels said: "I don't think I've ever seen scientists attacked so vehemently by people in Congress for their beliefs." He said that the report was "an attempt to chill debate".

Colin Macilwain

Berlin's universities face further cuts

Munich. Berlin's three universities reacted angrily last week to the city government's decision to cut their combined budget by DM150 million (US\$231 million) in the next five years, on top of the cut of DM600 million (\$924 million) imposed since 1993.

The government wants to reduce the universities' costs to DM2 billion by 2001 (see *Nature* 369, 431; 1994). Peter Radunski, Berlin's minister for science, believes that the cuts will be very damaging. He accepts the new cut, which has been caused by the city's near bankruptcy, with bad grace. "We have no choice," he says.

After years of battling with each other, the Humboldt, Free and Technical Universities have recently begun cooperating to find ways of coping with the funding crisis. Over the summer, as it became clear that new cuts were likely, they told Radunski that they would accept a reduction of DM50 million in return for an assurance of no further cuts before 2001.

The proposed agreement also included compromises on some controversial issues — for example, the universities would introduce entrance examinations and streamline their administrations.

Radunski now wants to offer the universities a less favourable deal: no cuts for four years in exchange for the DM150-million cut imposed last week. But the universities may not be ready to negotiate with him. Dieter Schumann, president of the Technical University, says: "The common basis for a contract has been destroyed because the new plans will be our ruin."

Quirin Schiermeier

Maths gaffe may add up to negative votes

Washington. George Brown (Democrat, California), a leading science advocate in the US Congress (see above), has come under heavy fire from his Republican opponent, Linda Wilde, after a television debate *faux pas* which could affect his success in defending a wafer-thin majority in his San Bernardino district.

At one point in the debate, the congressman said that he imagined that Wilde, "because she is a lady, is afraid of math". His opponent, whose father is a mathematician, subsequently went after Brown demanding further apologies — Brown had already apologised, before the debate was over — and using it to portray the 76-year-old liberal congress-

man as sexist and out-of-touch.

Brown later characterized his comment as a "misstatement", saying that he had been trying to underline the need encourage school teachers to overcome false stereotypes — such as the idea that girls shouldn't do mathematics. "My record on women's issues speaks for itself," he says.

In next week's election, Brown, a leading congressional liberal, is defending a district which he has held for sixteen consecutive terms in the face of demographic changes that make it look increasingly like Republican territory. "George is not good at telling jokes," sighs one of his campaign staff. **C. M.**

Election tonic for Japanese science funding

Tokyo. Japan's government research sector appears likely to win some more money with the expected return to power of the Liberal Democratic Party (LDP) after last week's general election. But there are unlikely to be any radical changes to the administration of science, despite talk by all political parties, during the election campaign, about the need to reform and reduce the size of government ministries and agencies.

The LDP increased its number of seats in the powerful lower house of the Diet (parliament) from 211 to 239 — 12 short of a majority. It seems set to take power alone, without its former coalition partners, the Social Democratic and Sakigake parties. Talks on maintaining the coalition government failed. Instead, the two former coalition partners, which lost many seats in the election, have agreed to continue to cooperate with the LDP on common goals.

But this will put the LDP in a weak position, and it will probably try to gain the support of independent Diet members. It may also be able to persuade some former LDP politicians in the New Frontier Party (Shinshintō) to return to the fold. It seems certain, however, that Ryutaro Hashimoto, the head of the LDP, will be re-elected as prime minister early next month when the Diet reconvenes.

The LDP places top priority on introducing a supplementary budget before the end of the year to spend ¥5,000 billion (US\$45 billion) to help boost the flagging economy. And last week Hashimoto singled out

science and technology and research and development as areas that should get some of the extra funds.

Although the other parties are lukewarm on the need for a supplementary budget,

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Hashimoto: regards science and technology as key areas to benefit from extra money.

there is broad cross-party support for giving extra money to science and technology. The supplementary budget would add to several others over the past few years that have pumped substantial new funds into government research.

Another beneficiary on the side of science is likely to be the little island of Okinawa at the southern end of Japan. Just before the election, Hashimoto reached an agreement with the governor of Okinawa to pump several billion yen of extra government money into the island in return for cooperation in keeping US military bases on the island. Some of the money is expected to go into new facilities for "tropical research".

Although the Social Democratic Party is

pushing for much greater reduction of US military facilities in Okinawa, it is unlikely to oppose the new funds.

Officials in government ministries and agencies will be happy to see the LDP back at the helm. In the lead-up to the election, many parties, including the LDP, were promising radical reform of the bureaucracy, which has been blamed for the crisis in the banking system and for the HIV contamination of haemophiliacs. The LDP, however, has traditionally maintained strong links with the administrative wing of government, and any reform it brings will be much less radical than anything proposed by the opposition parties.

But Akito Arima, president of the Institute of Physical and Chemical Research (RIKEN) and a leading campaigner for reform of government research, says there may still be serious discussion on reorganizing the roles of the key science-related ministries. These include the Ministry of International Trade and Industry, the Science and Technology Agency, the Ministry of Health and Welfare, and the Ministry of Education, Science, Sports and Culture.

In its weakened position without coalition partners, the LDP may be forced to listen to calls for reform from opposition parties, such as the new Democratic Party of Japan. One of this party's joint leaders is Naoto Kan, who, as minister of health in Hashimoto's government, exposed his ministry's role in the infection of haemophiliacs with HIV.

David Swinbanks

Biologists call on Brussels to bolster basic research

Munich. A panel of European biologists has urged the European Commission to ensure that basic research receives strong support in the commission's next five-year Framework research programme (FP5). In particular, the panel, which was chaired by François Gros, former director of the Pasteur Institute in Paris, says there is a need for more European-level support for genetics and neuroscience research.

The *ad hoc* panel, an informal group made up largely of members of the European Science and Technology Assembly (ESTA), an advisory body to the European Commission, was set up earlier this year to advise ESTA on the recommendations that the assembly should make to the commission about support for biology research under FP5.

It presented its conclusions about factors that the commission should consider in setting European priorities for biology to the assembly earlier this month.

The panel's report argues that it is important to include basic research in FP5, which runs from 1998 to 2002, because its

long-term survival is essential for the future of applied research, and it would be a "serious mistake to overlook the interdependence of advances in science and technology".

The biologists warn against rigidly applying the principle of subsidiarity, which prevents the commission from supporting action that can be supported at a national level.

A narrow interpretation of this principle, they say, would exclude European-level support for long-term basic research, as most of this is seen as a national responsibility.

Specific areas that should receive more support, they say, are the creation of scientific networks, allowing the building-up of international research teams, and the funding of large facilities such as major instruments, laboratories and databases.

Because of their size, scope and expense, such facilities should be accessible to scientists in all European countries. The group also recommends that FP5 should give the commission the opportunity to give

additional funds to existing national or European centres of excellence.

The report focuses on the advantages of European-level research in genetics and molecular biology, where the commission has already made considerable investment in infrastructure. It has approved two large grants to establish a European mouse-gene repository at Monterotondo near Rome (EMMA), and four supporting laboratories in France, Sweden, Portugal and Britain.

The repository will house a collection of mutant-mouse strains, which are widely used, for example, as experimental models of human diseases. The commission has also supported the development of the European Bioinformatics Institute (EBI) in Cambridge, a genetic database which is an outstation of the European Molecular Biology Laboratory.

The report says that this type of initiative requires a longer-term commitment than the commission is able to guarantee within its five-year Framework programmes, although it makes no suggestion about how this could be achieved.

Alison Abbott

Nuclear waste plans enter critical phase

London. Britain's newly privatized nuclear power industry is facing a critical few months as it contemplates how to dispose of up to 300,000 cubic metres of radioactive waste by early next century — and answer critics who claim that its plans are unsafe.

At stake is the outcome of a 66-day planning inquiry into an experimental underground rock laboratory below Sellafield, Cumbria, in northwest England. The industry's waste disposal company, Nirex, claims it needs the laboratory to assess the suitability of the rocks for a deep repository for low- and intermediate-level radioactive waste.

But Nirex needs to convince Cumbria County Council of the desirability of its plans. The council is opposed to the plans, and has already rejected one planning application for the rock laboratory. The council, with some scientists and environmental groups, says more basic research is needed. It is also concerned that the laboratory may be converted into part of the repository.

The final decision, based on an inspector's report of the inquiry, which ended on 1 February, is now awaited from John Gummer, secretary of state for the environment. If the decision goes against Nirex, its plans would be set back at least five years, which could prove costly. Perhaps more damagingly, Nirex could be faced with the prospect of being turned away by every local planning authority in the country.

To make matters worse, Nirex has been criticized by the Royal Society for holding back important data that need to be assessed by government regulators, and for not subjecting its science to independent peer review.

Critics of Nirex's approach also include the UK government's Radioactive Waste Management Advisory Committee

(RWMAC), which is compiling a report on Nirex and peer review, as well as the environmental groups Friends of the Earth (FoE) and Greenpeace.

The situation has resulted in a tense debate within Nirex. The company's scientists are unlikely to give their approval to a Sellafield repository unless they are convinced that the site is safe. But major shareholders are said to be becoming restless. The longer the issue is left unresolved, the greater the likelihood that they could lose potential overseas contracts.

Nirex has already spent £400 million (US\$636 million) on research aimed at assessing the safety of underground disposal, particularly after the repository's expected 50-year period of use. It has drilled 27 boreholes to investigate the geology and hydrogeology. A further £200 million will be spent on the rock laboratory.

Nirex's problems began in 1991, when it announced Sellafield as its proposed repository site from a short-list of 12 locations. According to the company, Sellafield was chosen using a technique known as multi-attribute decision analysis, in which each site is assessed by adding up scores assigned to technical, social and economic considerations. But a detailed explanation of this decision was not released immediately.

This was pointed out by the Royal Society in its report on radioactive waste disposal published in November 1994, which also acknowledged that the site was convenient, as most of the waste would come from the nearby Sellafield reprocessing plant of British Nuclear Fuels (BNFL). The Royal Society nevertheless says it retains an "open mind" as to the suitability of Sellafield.

Andrew Stirling, a research fellow at the Science Policy Research Unit at the Univer-

sity of Sussex, assessed Nirex's more complete account of its decision-making on behalf of Greenpeace. He says that he found inconsistencies in the methodology. For example, "long-term public safety was assigned a very low weighting". Nirex justified this on the grounds that the site would not have been shortlisted if its safety to the public had been in question.

Nirex's research plans, however, have been welcomed by many in the scientific community, including RWMAC and the Royal Society, with the extra caveats that the company must provide enough time to gather and assess data, and subject its science to independent peer review. Nirex has announced that it plans to deposit waste in the repository by 2012. But John Holmes, the agency's director of science, says this date is not set in stone and depends on the results of the scientific investigation.

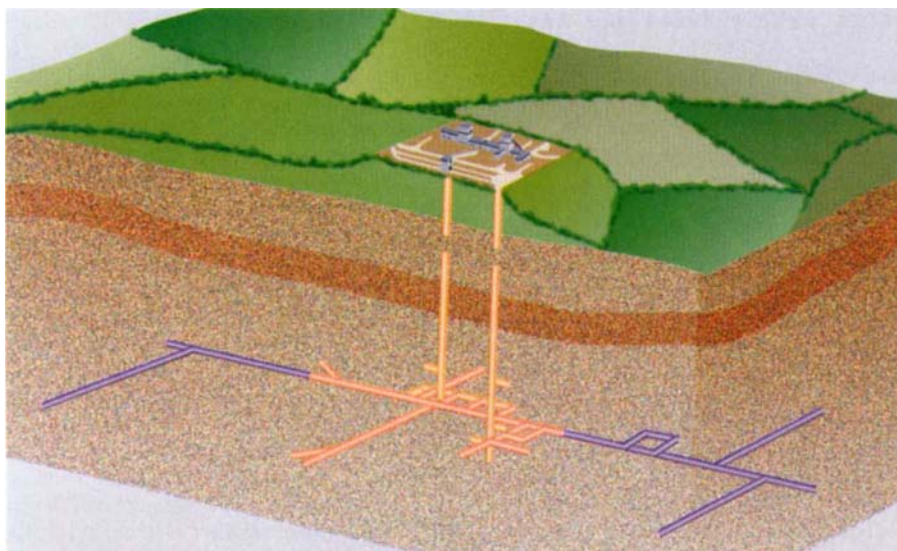
But others are unsure about Nirex's research plans. David Smythe, professor of geophysics at the University of Glasgow, was a member of a BNFL review panel assessing possible repository sites. Smythe says he was initially in favour of Sellafield, but changed his mind when the site's geology was found to be more complex than previously thought. He resigned in 1991, sensing that the nuclear industry was using the panel to "rubber stamp" the choice of Sellafield. "It was science in the service of politics."

Smythe is one of a number of observers who believe that Nirex's scientists are still caught up in the old 'secrecy culture' that Britain's nuclear industry is slowly beginning to shake off. One scientist who has worked extensively in radioactive waste disposal, but who declines to be identified, says Nirex scientists have continually to struggle to convince the company's management of the need to make research results public.

Rachel Western, a nuclear researcher for FoE, points out that Nirex's 'safety studies', normally published about 20 times a year, were reduced to a trickle during the rock laboratory inquiry.

Holmes says the reduced frequency of the studies is temporary as Nirex had to devote all available resources to the inquiry. He adds that the company publishes far more than comparable commercial organizations, such as British Coal. "But because we carry the nuclear label, there is this presumption that we are hiding something," Nirex, he says, plans to publish more often and to have some work reviewed by an outside panel of scientists.

But there remains no completely independent mechanism to assess the company's research. Nirex will need to apply to the government's Environment Agency for authorization to build the repository. But under recent government regulations — known as 'the polluter pays' — Nirex pays to have ►



Going underground: Nirex claims a rock laboratory (above) is needed to assess site safety. Critics contend that shaft-sinking could disturb the hydrogeology Nirex seeks to examine.

► its own research assessed.

This arrangement has already run into trouble. Almost two years after announcing its decision to construct a rock laboratory, Nirex pulled out of an arrangement with the government's pollution inspectors to fund an assessment of its waste disposal plans, midway through the study.

In a report, the inspectors noted that Nirex had delayed providing key information. The Royal Society, too, noted communication difficulties between the inspectors and Nirex. Holmes, however, says that particular assessment was geared to wards lodging a repository planning application in 1992, and ended "by mutual arrangement". Both organizations, he adds, are consulting over a fresh research assessment.

Miro Ivanovich, professor of geochemistry at the University of Reading, who has been involved in Nirex work, says future assessments must be made more transparent. He says Nirex should be asked to place its research funds under independent control. "I'm not criticizing the quality of what they [Nirex] do," he says. "But people must be convinced that there is no conspiracy."

Part of the problem stems from the shortage of independent hydrogeological expertise in the UK radioactive waste management community. Western says that both FoE and Cumbria County Council had difficulty in finding independent experts to review Nirex's evidence to the rock laboratory inquiry. "The majority of hydrogeologists working in radioactive waste management have had some contact with Nirex," says Jane Dotteridge, lecturer in hydrogeology at University College London and chair of the hydrogeology group of the Geological Society of London.

Dotteridge says that graduates from several new MSc courses in radioactive waste management could herald a change. But this is unlikely while Nirex remains the

Rock laboratory faces key questions

London. Before the seal is set on Sellafield in Cumbria, northwest England, as the home of Britain's first underground nuclear waste repository, Nirex will need to convince itself — and its critics — that the waste poses no danger to present and future generations.

The greatest potential threat is from radionuclides finding their way back to the surface long after the cement-filled steel drums containing the waste have been corroded by groundwater seeping into the repository. Nirex needs to assess how quickly water will enter the repository and corrode the containers.

One of the most difficult tasks is to predict the fracture pattern of the rocks, and then to calculate the speed and direction of the groundwater flow. Nirex's safety assessment also needs to consider the effects of gas deposits, earthquakes and climate change — an ice age is predicted to return to Cumbria within 100,000 years.

The organization has begun to address some of these issues by drilling deep boreholes near the planned repository. It now wants to build its controversial rock laboratory, consisting of two 5-metre-diameter shafts dug 790 and 920 metres into the ground, connected to a network of tunnels.

But critics believe there is more borehole work to be done. Opposition to the

scheme centres on concerns that Nirex has not fully considered the impact on the site's hydrogeological and hydrogeochemical conditions of constructing the laboratory.

The Radioactive Waste Management Advisory Committee (RWMAC) and Friends of the Earth have pointed out that Nirex must provide evidence that building the laboratory will not disturb the natural environment that it aims to monitor. John Holmes, Nirex's director of science, says the company has commissioned four scientists to assess this issue. He says their recommendations will be published and acted upon.

In addition, the Royal Society and RWMAC have suggested that Nirex's relatively short research timetable may not yield enough data for a proper safety assessment and should be extended. Nirex plans to apply for planning permission for the repository by 1998–99. But these organizations point out that accurate data will not be obtained until environmental conditions settle down, which could take up to three years after the laboratory's construction.

Holmes says that Nirex is committed to keeping the rock laboratory open for nearly 10 years, and will walk away from the site if the results show that a Sellafield repository, as currently proposed, is potentially unsafe. **E. M.**

dominant funder of research.

Holmes discounts the need for any additional external reviews of Nirex's work, however, and says there "is no shortage of alternative opinions". The public inquiry for the repository will, he says, provide a good

test of Nirex's arguments. He says the grilling that Nirex received at the inquiry for the rock laboratory from environmentalist groups should be ample evidence that "there is no chance that we can pull the wool over people's eyes". **Ehsan Masood**

US panel approves salt bed burial in New Mexico

Washington. The United States can safely bury intermediate nuclear waste in a salt bed under southern New Mexico, provided that future generations do not try to drill for oil through the dump, according to a panel of experts convened by the National Research Council (NRC), the operating arm of the National Academy of Sciences.

Their recommendation will increase the chances that the proposed Waste Isolation Pilot Plant (WIPP) will be constructed. The Department of Energy, which wants to build the dump in order to dispose of plutonium cuttings and other debris from the nuclear weapons programme, will formally apply this week to the Environmental Protection Agency (EPA) for permission to do so.

The panel, chaired by Charles Fairhurst, a mining professor at the University of Minnesota, found that, in the absence of human disturbance, the proposed repository

would successfully isolate the barrels of solid waste for at least ten thousand years. In the event of drilling or other human activity, the panel said, engineering methods are available that would minimize the chances of radioactive releases.

Plans to build the WIPP near Carlsbad, New Mexico, have been under consideration for almost twenty years, and experimental storage areas have been dug out of the salt 2,000 feet underground. The salt bed is an attractive medium for dumping because the salt is virtually impermeable to water, but will gradually close up around the waste and seal it, like a viscous fluid.

WIPP would store up to 137,000 cubic metres of the solid waste, held in 55-gallon steel drums. The total radiation inventory in the repository would be a thousand times less than at the proposed store for spent nuclear fuel at Yucca Mountain, Nevada.

Unlike Yucca Mountain, WIPP has the support of many politicians in its home state.

"In the undisturbed condition, we see no credible scenario for the release of radionuclides" from the proposed dump, says Fairhurst, pointing that the situation is different in other countries which have abandoned the idea of repositories.

The panel noted that EPA regulations assume that oil and gas drilling activity will take place over the next 10,000 years at the rate now prevalent in the surrounding area. This assumption is "highly conjectural and lacks scientific foundation," Fairhurst says.

The panel suggests that the EPA should reconsider its assumptions about drilling activity, and that the Department of Energy should continue work on techniques to increase WIPP's resilience to human disturbance, the nature of which, it concedes, is impossible to predict. **Colin Macilwain**

Papal confession: Darwin was right about evolution

Munich. Pope John Paul II has acknowledged the existence of evolution, nearly 150 years after Charles Darwin introduced the world to the idea that humans may not date back to God's seven days of creation.

In a statement last week to the Pontifical Academy of Sciences, an international body of about 80 prominent scientists, the Pope accepted that the overwhelming volume of experimental evidence in support of evolution could no longer be ignored. "Observational sciences describe and measure with ever increasing precision the multiple manifestations of life and inscribe them in an [evolutionary] timescale," he said.

Darwin had never been subject to the prolonged vilification that Galileo Galilei suffered from the Roman Catholic Church



because of his rejection of Ptolemy in favour of Copernicus. Galileo was rehabilitated by John Paul II four years ago.

The strongest message from the Vatican on the theory of evolution had been in a encyclical letter issued by Pope Pius XII in 1950. The letter, called *Humani Generis (Of the Human Species)*, warned of the threat Darwinism posed to central tenets of the Catholic faith. But it said that evolutionary theory was generally acceptable, provided that it was presented only as a hypothesis and not as a doctrine.

Pope John Paul II has now indicated that the Catholic Church is ready formally to accept scientific evidence that evolution is more than just a hypothesis — something that has been accepted in practice by most Catholic theologians for many years.

But, in his statement to the academy, the Pope drew a firm line between the material and the spiritual. He said that it is acceptable to believe that "the human body originates from living matter which predates it", but not that "the spirit is also a product of matter". The latter would lead to an unresolvable conflict between science and faith.

Alison Abbott

Clinton woos women's votes with cancer genetics funds

Washington. Ten days before next week's presidential election, in which the votes of women are expected to play a decisive role, President Bill Clinton announced last Sunday that he will redirect US\$30 million to breast cancer genetics research in 1997.

Exactly \$20 million will come from a \$100-million breast cancer research programme run by the US Army. That project has come under fire from some scientists, who claim it has ignored the scientific merits of grant proposals in pursuit of political ends (see *Nature* 383, 113; 1996). It will be instructed to devote \$20 million to breast cancer genetics grants "in collaboration with" the National Institutes of Health (NIH).

The president, whose mother died of breast cancer last year, made the announcement standing in the White House rose garden, surrounded by breast cancer survivors. He said the money will "support and expand breast cancer genetic research at hospitals, universities and laboratories across America". He called the step "a major increase in breast cancer genetic research".

The NIH spent about \$60 million on breast cancer genetics research in 1996, according to the office of director Harold Varmus. "What we're doing [with the new money] is putting greater emphasis on research on the genetics of breast cancer," says Anne Thomas, a spokeswoman for Varmus. "We want to take advantage of the progress that's been made in this area."

The new money has already been committed to biomedical research in the 1997 budget, which Clinton signed into law in September. It will be redirected within existing programmes. In addition to the money in the US Army programme, \$10

million will come from the NIH through the National Cancer Institute (NCI), the National Center for Human Genome Research, the National Institute of Environmental Health Sciences, Varmus's office, and elsewhere. Most of the money will be spent on investigator-initiated grants to study breast cancer genetics.

Reactions among breast cancer advocacy groups has been mixed. "We are always thrilled when there is an effort to recognize breast cancer as the significant disease to our society that it is," says Mary Jo Ellis Kahn of the National Breast Cancer Coalition. But she added that her group fears that the infusion of money might encourage women to rush out and seek testing for genetic predisposition to breast cancer — an area "fraught with incredible danger" to women because no comprehensive federal law bars employment and insurance discrimination based on the results of genetic tests.

Others charged the president with transparent politicking. Clinton's announcement, less than two weeks before voters decide on whether to grant him a second term, "illustrates the problems inherent in government funding scientific research", says Michael Tanner, a health policy analyst at the Cato Institute, a libertarian Washington think-tank.

The president has made a priority of appealing to women, a key swing-voting bloc, enacting laws ranging from unpaid family leave to longer hospital stays when women give birth.

The NIH expects to spend about \$430 million of its total budget of \$12.7 billion on breast cancer research in 1997.

Meredith Wadman

BSE link prompts German vote on sheep

London. The German cabinet is to vote next week on a proposal from its federal health ministry to ban the importation from Britain and France of certain sheep parts — in particular brains, spinal cords and eyeballs — out of concern that they could have become infected with bovine spongiform encephalopathy (BSE).

The decision comes soon after last week's announcement of the first scientific evidence, produced by a team headed by John Collinge of Imperial College Medical School in London, of a direct link between BSE and a new strain of Creutzfeldt-Jakob disease (CJD) in humans (see *Nature* 383, 685-690; 1996).

Speaking at a press conference in London last week, Collinge said that one of the possible applications of his discovery of

biochemical evidence linking the two diseases directly was that it might also be used to look for the emergence of BSE in other animals.

In particular, it was now possible, at least in theory, to distinguish sheep that could have been infected with BSE from those suffering from one of the many different strains of scrapie, another prion disease which is thought by some to be at the origin of BSE itself. "It is possible that BSE may have gone back into sheep," says Collinge.

British officials point out that it has been compulsory to destroy the heads of both sheep and goats at the point of slaughter as a precautionary measure since last summer. The main issue of contention at present is whether the spinal cord should be left in carcasses that are exported. □

Space agency rejects proposal to split science directorate

Munich. The science programme of the European Space Agency (ESA) is to retain its own identity in a revised management structure for the agency, following the rejection of plans that would have separated responsibility for planning and implementation into two separate directorates.

Scientists had argued that chopping up the science programme would have reduced its efficiency (see *Nature* **381**, 353; 1996), and Jean-Marie Luton, ESA's director general, told an ESA council meeting last week that the science directorate would remain unchanged in the new management scheme, which will be implemented next year.

The science directorate in addition will take on responsibility for the science part of ESA's Earth observation programmes. But such programmes will not share the general science programme's privilege of mandatory financing, and will remain on an optional menu for ESA's 14 member states. □

Europe and Russia plan pact

Moscow. Edith Cresson, the European Union's commissioner of research, and Vladimir Fortov, newly appointed chairman of the Russian State Committee on Science and Technologies, have signed a joint declaration stating that the Russian Federation and the European Union intend to conclude a treaty on cooperation in science and technologies.

Cresson said that Russia heads the list of states of the former Soviet Union whose research efforts are being supported by the European Commission. For example, four groups of Russian scientists are participating in the Esprit programme concerned with

information technologies. She said that a major goal of the European Commission is to secure for Russian scientists "the normal conditions of work at home", so that they are not tempted to leave the country. But she added: "Russia should definitely name its priorities in scientific research". □

UK resistance to EU research

London. Funds received from the European Commission in Brussels can have a disproportionate impact on Britain's research priorities because, under UK practices, funds 'attributed' to European-level research projects are deducted from the budgets of government departments, according to a report published earlier this month by the Parliamentary Office of Science and Technology (POST).

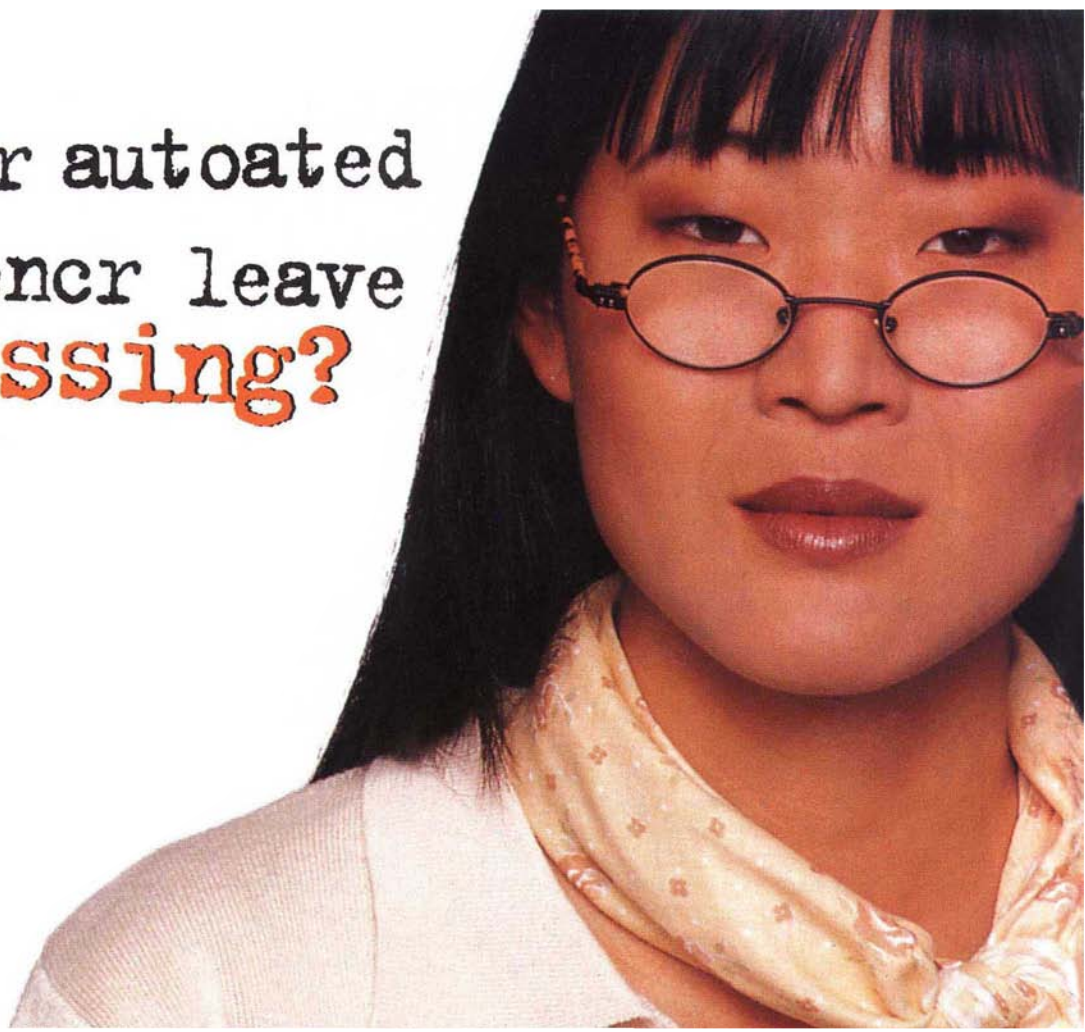
According to the POST report, the result of this procedure can be a reluctance on the part of such departments to take part in research programmes of the European Union, out of fear that departmental funds may be cut, and a resistance — because of the need to observe departmental accountability — to interdisciplinary projects. □

Lab mismanagement costs dear

San Francisco. The University of California has agreed to pay \$2.7 million to the US government to settle overcharges and mishandling of funds on federal projects at Lawrence Livermore Laboratory from 1990 to 1993. The payment, which includes about \$1.5 million to cover overcharges plus \$1.2 million in damages, is the largest settlement made during the university's management of the laboratory.

The laboratory acknowledged that it had charged some costs to other project sponsors, made unauthorized loans from one project to another, and neglected to return unused funds. But it also undercharged some government agencies by comparable amounts, and the US government did not therefore lose money overall.

Does yur autoated
DNA seqencer leave
u guessing?



Furthermore, according to a laboratory spokesman, there was no personal gain by any employee as a result of the mismanagement. □

Robotic telescope on-line

San Francisco. The University of California, Santa Cruz, geared up last week for the full automation of a 30-inch telescope at Lick Observatory that university officials claim will be the most sensitive fully robotic telescope in the world.

Built at the University of California at Berkeley under the leadership of the astronomer Alex Filippenko, the Katzman Automatic Imaging Telescope (KAIT) is unusual because of a complex tracking system that can remain fixed on a particular area of the sky for hours. Complex software also allows the orchestration of a night's observations, based on target requests sent through e-mail. The instrument took seven years to build, and Filippenko plans to use it mainly to observe exploding stars or supernovas. □

AIDS in Europe 'stabilizes'

Paris. The AIDS epidemic in Europe "seems overall to have stabilized", according to a report from the Paris-based European Centre for the Epidemiological Monitoring of Aids. But the centre emphasizes that this finding must be interpreted with care, given the geographical heterogeneity of the epidemic within the region.

A total of 13,310 cases of AIDS were recorded in the first six months of this year among the 45 countries making up the World Health Organization's European grouping — making a total of 174,260 cases in this region since the beginning of the epidemic. Incidence has stabilized in northern countries over the past two to three years, mainly due to the plateauing of cases of homosexual transmission. But there is no sign of a stabilization of the epidemic in southern countries such as Spain and Italy, mainly because of new cases arising among drug addicts using tainted syringes. □

Europe plugs away in vain

Brussels. After five years of strenuous effort, member states of the European Union have failed to agree on a single common design for power plugs and sockets. Six different plug designs are used in European states, varying from the three square pins used in Great Britain (with a fuse incorporated in the plug) to the two circular pins used in much of the rest of the continent. Despite the difficulties this causes to travellers and to manufacturers, Cenelec, the electrical standards body, last week rejected a plan to agree on a single 'Europlug'. Klaus Kinkel, Germany's foreign minister, said it "must not be a model for monetary union or enlargement of the European Union". □

Seveso dioxin levels 'still too high'

London. The environmental group Greenpeace claimed last week that it had discovered levels of dioxin in the area around Seveso, northern Italy, where a chemical explosion took place 20 years ago, that are ten times higher than those that the Italian government's agriculture advisory committee considers safe for food production.

The group claims the levels of dioxin in the Seveso region remain close to those recorded in 1982, alleging that in the intervening period, "nothing has been done to solve the problem of dioxin contamination in the areas of the accident". The 1976 explosion took place in a factory belonging to an affiliate of the Swiss pharmaceutical company, Roche. It led to the deaths of farm animals in the region, and to an outbreak of skin diseases among local residents. □

Science as hanging judge

Washington. Scientists in the United States face a fresh challenge if Wayne Allard, the Republican candidate, wins a tight Senate race in Colorado during next week's federal elections. Public hangings should resume, he says, "if scientists proved they deter crime". □

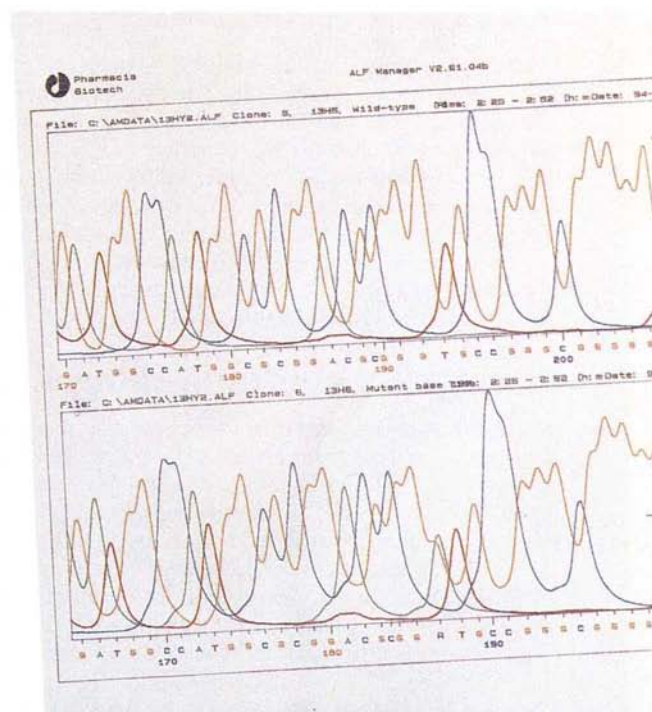
Does your automated DNA sequencer leave you guessing? If so, chances are it's primarily designed for high throughput sequencing. Why be uncertain of your sequencer's accuracy, when ALFexpress™ is providing researchers with the full genetic stories of their DNA.

ALFexpress: for more accurate readings

In the largest clinical study using automated DNA sequencing, the technology behind ALFexpress proved exceptional (see caption). That's one of many examples of ALFexpress offering unrivalled accuracy during automated confirmatory sequencing. Further, its readings are so accurate that ALFexpress can unambiguously identify heterozygous point mutations—as proven in many clinical research applications, such as analysis of tumor genes and high-resolution HLA typing. What's more, Pharmacia Biotech has dedicated software programs to support these applications.

For the full story, call us: 1 (800) 526-3593 from the U.S.; +81 (0)3 3492 6949 from Japan; or +46 (0)18 16 50 00 from Europe and the rest of the world. Or visit us on the Internet: <http://www.biotech.pharmacia.se>.

Pharmacia Biotech
Uppsala, Sweden. (And the rest of the world)



The p53 gene from 316 breast cancer patients was sequenced using ALF automated sequencing technology. (Bergh J., Norberg, T., Sjögren, S., Lindgren A., Holmberg, L. "Complete Sequencing of the p53 Gene..." *Nature Medicine* 1995; 10:1029-1034.)



Plants condemned as pesticides

SIR — Meredith Wadman relates only part of the controversy over a new US Environmental Protection Agency (EPA) policy (*Nature* 382, 485; 1996). The EPA has in its regulatory crosshairs a promising new area of biotechnology — solely because it is biotechnology. The agency would regulate garden and crop plants as though they were pesticides. Case-by-case regulatory review would be required only when these familiar plants are genetically modified using recombinant-DNA methods to enhance their pest-resistance. Under the new EPA policy, biotechnology-improved varieties of corn, cotton, marigolds and so on are “pesticidal plants”. They would not only be regulated more stringently than other plants but also more stringently than chemicals similar to DDT or sarin. US pesticide regulations exempt small-scale field trials (less than 10 acres) of pesticides — whether chemicals or organisms — but this research exemption does not apply to rDNA-manipulated plants. Moreover, the regulations would require rDNA-improved plants to be labelled as “pesticides”.

Eleven scientific societies representing more than 80,000 biologists and food professionals published a comprehensive report condemning the EPA's proposal. The report emphasizes that the safety of a new substance synthesized by a plant depends on the biological actions of the substance and where it is expressed in the plant, rather than the fact that it is intended to protect against a plant pest, or on the use of certain genetic techniques.

The report warns that the EPA policy would have negative consequences, including an increased regulatory burden on research and development of pest-resistant crop varieties, the growth of bureaucracy, sustained need for chemical pesticides and competitive disadvantage to those using the new biotechnology.

Wadman points out that the major US industry trade association — the Biotechnology Industry Organization (BIO) — has joined anti-biotechnology groups such as the Environmental Defense Fund “in opposing such conclusions”. In fact, BIO and Monsanto, a company that dominates the association, have not only continued to support EPA's policy in the face of incontrovertible scientific and economic arguments against it, but have actually tried to persuade some of the societies to disavow the report. Alan Goldhammer of BIO and Monsanto's chief Washington lobbyist, Chester Dickerson, have put pressure on representatives of the agronomy, crop science, microbiology, food technology and weed science societies that collaborated on the report.

The EPA policy creates new barriers for

smaller, cash-strapped companies. It could virtually extinguish university research on the nature of pest-plant interactions, which is essential for developing alternatives to agricultural chemicals. The large companies that control BIO's positions on agricultural issues are among the world's largest producers of agricultural chemicals — Monsanto's sales of crop- and lawn-protection products in 1995 were \$2.47 billion, for example. These companies therefore have a vested interest in slowing the development of competing, potentially superior products.

At a fundamental level, this collusive, anti-competitive behaviour corrupts one of the principles of the free market — that entities are effectively free to enter into and carry on business transactions. The cynical EPA/BIO/Monsanto strategy makes losers of the research community, entrepreneurial biotechnology companies and consumers.

Henry I. Miller

*Hoover Institution & Institute
for International Studies,
Stanford University,
Stanford, California 94305-6010, USA
e-mail: miller@hoover.stanford.edu*

Did the Romans exploit uranium?

SIR — In my recent book *Uranium Glass*, I discussed some Roman mosaic glass of the first century AD that was excavated early in this century near Naples by the Oxford archaeologist R. T. Gunther, who reported that some of its tesserae were made from uranium glass.

Fission-track dating experiments conducted in the United States and Germany showed them to be no older than the beginning of this century. Moreover, US and German teams scanned the site with a radiation detector, but failed to find anything of interest. Doubts were consequently expressed as to whether the Romans really had exploited uranium; such a view appeared in the review of my book (in *Nature* 379, 34; 1996).

There are, however, good reasons for believing that the Romans probably did exploit uranium. Gunther found a uranium-oxide content of 1.25% by means of chemical analysis (most Roman glass contains only 0.0005%), and he tried to replicate the glass using these results, both with and without added uranium. There were now three kinds of glass: the original Roman glass, the replicated uranium glass and the replicated non-uranium glass.

In 1948, Earle R. Caley reviewed Gunther's report and concluded that the uranium content ought to have been shown

as 1.5% rather than 1.25%. In 1963, Franz Kirchheimer analysed specimens of the mosaic glass in Oxford and observed a uranium oxide content of 1.6%, a value close to Caley's 1.5%. His analysis also indicated a lead content of about 1%, about which Gunther was silent. If Kirchheimer's experiments had involved only the replicated uranium glass specimens, he would not have discovered a uranium oxide content significantly larger than 1.25% (as he did), or such an appreciable content of lead. These facts strongly suggest the existence of Roman uranium glass.

The reasons for doubts are not all that strong either. The age of the specimens (less than 100 years) was probably obtained because they had once been melted. Any glass once heated shows only the age after the heating. The unsuccessful surveys in Naples merely suggest that the mosaic glass was unusual for the Roman period.

In 1977, Kirchheimer wrote a detailed case for accepting the existence of Roman uranium glass. He emphasized the significant lead content of a specimen preserved in a test-tube labelled “Oxford Roman glass with Uranium”. Although he was aware of the critical arguments, Kirchheimer stood by the existence of Roman uranium glass.

Ken Tomabechi

*4-15-14, Nakahara,
Mitaka-shi,
Tokyo 181, Japan*

Dousing the flame

SIR — In recent years, the Philip Morris Prize for Scientific and Technological Research has been advertised in leading Italian newspapers, and three government bodies have given it their patronage — the National Research Council (CNR), the Ministry of University and Research and the Italian Board of Energy.

During the same period, advertising by tobacco companies has evolved from showing an image of satisfied smokers to highlighting the association of tobacco with positive social values. In Western countries, scientific research is perceived as one of the foundations of welfare; it is therefore potentially rewarding for the Philip Morris Company to link tobacco with such values.

But scientific research should not be associated with smoking, the leading cause of ill-health and death in the Western world. The three government departments concerned should dissociate themselves from any future Philip Morris initiatives.

Paolo Crosignani

Franco Berrino

Franco Rilke

*Istituto Nazionale per lo Studio
e la Cura dei Tumori,
Via Venezian, 1,
I-20133 Milano, Italy*

Fishy fragments tip the scales

Philippe Janvier

The discovery of 510-million-year-old skeletal fragments that are thought to belong to an early armoured fish — probably one of the earliest-known vertebrates — may change accepted ideas about the evolution of the vertebrate skeleton.

VERTEBRATES have two skeletons: the internal or endoskeleton, and the external or exoskeleton, which is thought to have appeared later in the evolutionary history of this group. The endoskeleton was initially cartilaginous, and has become calcified or strengthened by bone over time. But the early evolutionary history of the exoskeleton is much more complex, both in its tissue structure and in its organization. In particular, the fossil 'armoured' jawless vertebrates, which were hitherto known to have lived from about 470 to 375 Myr ago (see figure), display an amazing diversity of exoskeletal structures that have long puzzled palaeontologists.

Now the discovery of 510-Myr-old armour fragments from Cambrian rocks in Australia, reported by Young *et al.*¹ on page 810, has increased this diversity still further. They suggest that these fragments belong to one of the earliest vertebrates, casting doubt on the theory that the primitive vertebrate exoskeleton was made up of minute, independent scales, and indicating instead that it may have consisted of relatively large plates.

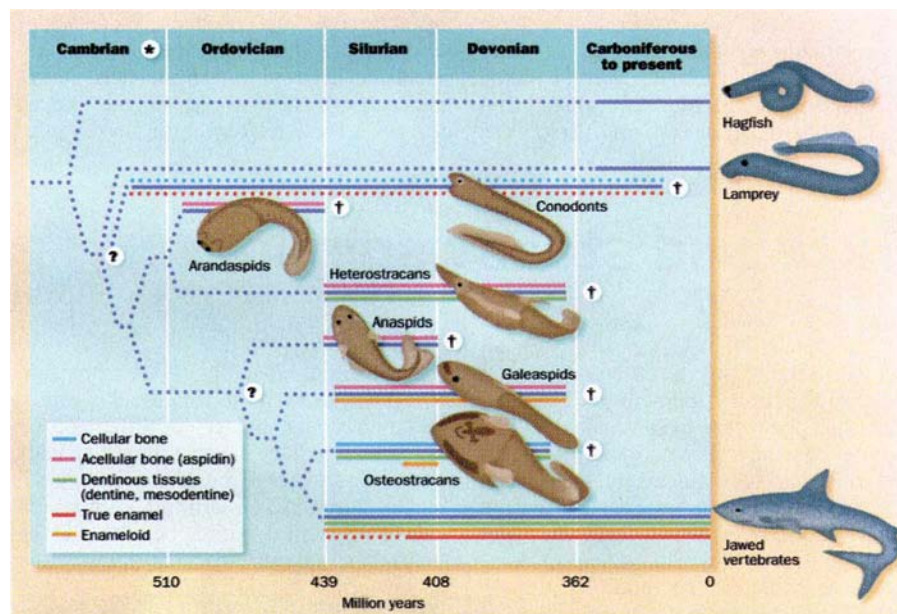
Most of the early vertebrates had a cartilaginous endoskeleton which, unless partly calcified, is not preserved in fossils and is therefore rarely informative. In contrast, much information regarding the morphology, phylogenetic relationships and evolution of these fossils can be extracted from their exoskeleton. However, the information provided by the morphology of the exoskeleton is rather limited, and palaeontologists have long tried to gain further information by examining the microstructure of the exoskeletal hard tissues of these early fossils (that is, whether they contain cavities for bone cells or tooth cells, and how their mineral components are organized). This approach has met with remarkable success over the past few years thanks to new observation techniques².

The method that has been used is to study early-vertebrate hard-tissue structure within the framework of the classical present-day vertebrate hard tissues — that is, bone with or without cells, various types of dentine (ivory), true enamel and enamel-like tissues (enameloids). This has worked, to a certain extent, for some groups of early vertebrates, such as heterostracans, osteostracans (see figure) or thelodonts, all of which possess

dentine in their exoskeleton. The osteostracans also have typical cellular bone (like higher vertebrates), and these classical exoskeletal tissues can be found from as far back as Late Ordovician times² (450 Myr ago).

The situation became more complicated when other groups from the Silurian and Devonian periods (430–365 Myr ago) were shown to have neither cellular bone

Palaeozoic period (ranging from about 520 to 450 Myr ago) have been studied. Many of these fragments were thought to be derived from vertebrates on the basis that they were made up of calcium phosphate — like bone and teeth. Some of these remains turned out to belong to crustaceans, lampshells (brachiopods), or extinct invertebrate groups that could produce a carapace or shell that vaguely re-



Distribution of the major groups of living and fossil vertebrates through time (solid mauve lines: actual distribution; dashed mauve lines: inferred distribution and relationships, according to one of the current theories). The distribution of the classical exoskeletal tissues is indicated by coloured lines (solid: actual distribution; dashed: debated distribution). The asterisk indicates the position in the geological timescale of the presumed vertebrate exoskeleton fragment described by Young *et al.*¹.

nor dentine (for example, anaspids or galeaspids; see figure). Instead, they had a lamellar layer of acellular bone, sometimes overlain by an enamel-like layer³. So began heated debates as to whether cellular or acellular bone and dentine or enamel were primitive. The criteria that were used to define 'primitiveness' of the various hard-tissue types were their great geological age, their widespread distribution, or their distribution in the framework of phylogenetic analyses that were based on many other characteristics.

To determine which kind of hard-tissue structure appeared first in the vertebrate exoskeleton and to find more about the diversity of early exoskeletons, all kinds of fragmentary remains from the Early

vertebrate bone in that it was made of calcium phosphate. But others remained more 'fishy' to the noses of palaeontologists, who tentatively referred to them as primitive vertebrates⁴.

Some people were sceptical about such claims — they believed that it was pointless to consider this kind of material while no complete specimen was known⁵. However, stories of this kind are not new in vertebrate palaeontology. For example, the Devonian heterostracan *Pteraspis* had acellular exoskeletal plates that were thought, in 1847, to be squid shells by the German palaeontologist Kner, and in 1858 were described as fish bones by T. H. Huxley. The clue came in 1863 when Huxley's protégé, the

famous zoologist Sir E. R. Lankester, found *Pteraspis* plates associated with a scaly tail. This shows that, despite uncertainty about their nature, it is important that palaeontologists are aware of these enigmatic bone-like remains.

Young *et al.*¹ did just this. The new fragment that they describe is particularly interesting because it bears rounded projections, or tubercles, that strikingly resemble those of arandaspids, a group of accepted jawless vertebrates from the Ordovician period⁶ (see figure). Although the tubercles in the Australian fragment are made up of an enamel-like tissue, whereas in arandaspids they consist of presumably acellular bone, in both cases there is no evidence of dentine. Some kind of enamel-like tissue, produced by the superficial layer of the skin, may be one of the first exoskeletal hard tissues to have appeared in vertebrate evolution.

Young *et al.* went on to ask how the first vertebrate exoskeleton might have been organized. The discovery of complete conodonts (see figure), with a naked body and powerful 'teeth', suggested that some kinds of teeth appeared first, even

if not associated with real jaws^{1,7}. Then, according to current theories, the rest of the exoskeleton developed over the body in the form of minute denticles (more or less like the scales of the present-day sharks). These denticles would later have fused to form large plates and scales. Strangely, no such denticle occurs among the earliest (Early Ordovician) vertebrate remains (either presumed or duly accepted). All of these remains are already in the form of large elements, or armour plates. This new Australian specimen, which is slightly older than the earliest-known undisputed vertebrate remains (arandaspids), confirms this observation and, as the authors suggest, may tell us that the vertebrate exoskeleton first developed in the form of large, mineralized units that were pervaded by canals for a sensory system.

Whatever the surprise that we may have when the owner of this debris turns up as a complete specimen, the work of Young *et al.*, like other papers on early remains of debated vertebrate derivation⁴, illustrates a new trend in the search for the oldest evidence for vertebrates.

Palaeontologists now dare venture into describing and interpreting this kind of Early Palaeozoic fossil, which looks 'fishy' in some respects but whose microscopic structure is sometimes at odds with the currently accepted structure of the vertebrate exoskeleton. This is certainly a risky task, as some of these remains may turn out to belong to the shell or carapace of some unknown invertebrate animal. But it is a risk worth taking in order to increase the probability of finding a clue to the still unresolved problem of the early evolution of the vertebrate skeleton. □

Philippe Janvier is in the Laboratoire de Paléontologie (URA 12 of the CNRS), Muséum National d'Histoire Naturelle, 8, rue Buffon, 75005 Paris, France.

1. Young, G. C., Karatajute-Talimaa, V. N. & Smith, M. M. *Nature* **383**, 810–812 (1996).
2. Smith, M. M., Sansom, I. J. & Smith, M. P. *Geobios. Mem. Spec.* **19**, 65–70 (1995).
3. Janvier, P. *C.R. Acad. Sci.* **310**, 655–659 (1990).
4. Smith, M. P. & Sansom, I. J. *Geobios. Mem. Spec.* **19**, 61–63 (1995).
5. Ørvig, T. *Zool. Scripta* **18**, 427–446 (1989).
6. Gagnier, P. Y. *Ann. Paleontol.* **79**, 19–69 (1993).
7. Purnell, M. A. *Nature* **374**, 798–800 (1995).

GALACTIC DYNAMICS

Slow and steady spirals

Ellen G. Zweibel

SPIRAL nebulae were detected in the eighteenth century, long before it became clear that they were distant galaxies. The magnetic spiral arms reported in this journal in January¹ are much subtler features, but their nature may already have been deduced. On page 800 of this issue², Fan and Lou propose an explanation for these magnetic arms that has ramifications for the general theory of spiral structure in galaxies.

In the current view³, a spiral galaxy is an extended spheroid, containing most of the galactic mass but little luminosity, within which is embedded a rotating disk of stars and gas that is less massive but more luminous. The spiral arms do not contain much more material than the rest of the disk, but they are disproportionately bright because of the highly luminous, newly formed young stars concentrated along them.

Because the inner parts of a galactic disk rotate more quickly than the outer parts, any ordinary material arms would get rapidly wound up, and the spiral structure would be lost. For this reason, the theory that gained favour in the 1960s explained spiral structure in terms of density

waves⁴, rigidly rotating with a pattern speed slower than the rotational speed of the underlying galaxy. Although it is not clear what causes them, density waves, and in some galaxies gravitational perturbations by an orbiting companion galaxy or central bar, appear to be the main ingredients for a satisfactory

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Arms out of step. Slow density waves may explain why the spiral galaxy NGC6946 has magnetic arms in between the luminous ones.

theory of spiral structure in galaxies⁵.

Gravitational forces dominate all of these mechanisms. The interstellar gas in galactic disks is too diffuse to contribute appreciably to the gravitational field, but

it gets strongly compressed by the slight density variations of spiral structure because the gas flow relative to the spiral arms is supersonic. The compression triggers small-scale gravitational instabilities which lead to a high rate of star formation, explaining the dramatic optical appearance of spiral galaxies⁶.

The interstellar magnetic field is strongly compressed along with the gas as it is carried through spiral arms⁷. That increases the field strength and energizes cosmic-ray electrons gyrating about the magnetic field lines, increasing the intensity of the radio-frequency synchrotron radiation that the electrons emit, and thus producing observable magnetic arms in about the same place as the optical arms.

So the detection of narrow magnetic arms *between* the optical arms of the galaxy NGC6946, and not associated with any known gas-dynamical feature, was surprising¹. Fan and Lou can explain the arms using magnetic forces; their calculation is based on magnetohydrodynamic wave theory combined with density-wave theory. Magnetic forces are anisotropic (direction-dependent), which leads to multiple wave modes, much as anisotropic elastic solids have more than one mode of oscillation. In the case of a relatively weak magnetic field, the so-called fast mode corresponds to the standard spiral density wave, for which magnetic field and density

fluctuations are in phase. But Fan and Lou have pointed out that there is also a slow-mode wave branch. The magnetic and density fluctuations in a slow-mode spiral wave are about 90° out of phase, so the magnetic and density maxima are interleaved rather than coincident.

Because the slow waves nearly co-rotate with the gas, the spiral pattern is confined to a radial range corresponding to a narrow range of rotational velocities. In other words, slow-mode spiral patterns can exist only in regions of nearly rigid galactic rotation. If NGC6946 has a large region of nearly rigid rotation, that might explain why the slow mode dominates.

We have only begun to explore the implications of this provocative explanation for magnetic arms. Nearly three decades ago, calculations of interstellar gas flow through classical (fast) density waves lent credibility to density-wave theory by providing detailed explanations of optical spiral structure⁶. Similar calculations for gas flow through slow spiral waves could be compared with observations to provide tests of the slow-mode magnetic arm theory, although the small relative velocity between the gas and the wave means that the gas response is probably weaker than for fast spirals. There is also the larger question of how slow-mode waves are coupled to the known mechanisms that excite spiral structure in galaxies. That is imperfectly understood even for the relatively well-studied fast density waves⁵.

On the observational side, it is essential to search for more examples of the magnetic arm phenomenon, and to establish correlations between the presence of magnetic arms and properties of the underlying galaxy. The main prediction of slow-mode density-wave theory at present is that the galaxy must rotate nearly uniformly in the region where the pattern is seen. This prediction could be tested with a larger sample. At the same time, detailed study of the properties of the known magnetic arms (in NGC6946 and three other, less well-established cases) could provide further tests of the slow-wave model, particularly if any accompanying dynamical features are seen in the gas-density distribution. □

Ellen G. Zweibel is in the Joint Institute for Laboratory Astrophysics and the Department of Astrophysical, Planetary and Atmospheric Sciences, University of Colorado, Boulder, Colorado 80309, USA.

1. Beck, R. & Hoernes, P. *Nature* **379**, 47–49 (1996).
2. Fan, Z. & Lou, Y.-Q. *Nature* **383**, 800–802 (1996).
3. Mihalas, D. & Binney, J. *Galactic Astronomy* 2nd edn (Freeman, San Francisco, 1981).
4. Lin, C. C. & Shu, F. H. *Astrophys J.* **140**, 646–655 (1964).
5. Binney, J. & Tremaine, S. *Galactic Dynamics* (Princeton Univ. Press, NJ, 1987).
6. Shu, F. H. et al. *Astrophys J.* **173**, 557–592 (1972).
7. Roberts, W. W. & Yuan, C. *Astrophys J.* **161**, 887–902 (1970).

How does morphine work?

Leslie L. Iversen

MORPHINE — the main pharmacologically active alkaloid of the opium poppy — has been used in medicine for at least 4,000 years, and it remains one of the most important and powerful medicines for the treatment of severe pain. But the euphoriant and addictive properties of morphine, and particularly its synthetic derivative heroin, also mean that the abuse of these drugs represents a major medical and social problem in many developed countries.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Opium poppies (*Papaver somniferum*) from which morphine is extracted, and (inset) the chemical structure of morphine, whose mechanism of action Matthes *et al.*¹ have gone some way towards elucidating.

Now, on page 819, findings by Matthes and colleagues¹ shed new light on this old drug. They show that mice lacking the μ -opioid receptor (MOR) do not show any of the normal responses to morphine, and these mice do not become dependent on morphine after repeated treatment with the drug. These important findings point to the MOR as the principal mediator of the pharmacological effects of morphine.

Since the 1970s, a large research effort has been devoted to understanding how morphine and related opiate drugs work in the central nervous system. From pharmacological studies it is clear that morphine and its derivatives act on specific receptors that recognize the drugs in a stereochemically selective manner². Pert and Snyder³ were the first to show the existence of stereoselective opiate receptor-binding sites in fragments from rat brain membranes, offering for the first time a biochemical approach to opiate receptor studies.

Another major advance was the discovery of a family of small peptides that occur naturally in the brain and act as the normal ligands for opiate receptors — the enkephalins⁴, β -endorphin and the dynorphins⁵. Kosterlitz and Paterson then concluded from a detailed pharmacologi-

cal analysis that the effects of opiates are mediated by not one, but three, different opiate-receptor subtypes, termed μ , δ and κ (ref. 6). This conclusion was confirmed by the identification and cloning of the three mammalian genes that encode these receptor molecules⁷.

Morphine can act, to some extent, on all three of the opiate-receptor subtypes. However, the findings of Matthes *et al.* indicate that the MOR is by far the most important¹. Using embryonic stem-cell gene-knockout technology, they bred a strain of mouse that lacked any detectable binding of MOR-selective, radioactively labelled ligands in brain homogenates. On treatment with morphine, these mice did not show any of the normal reductions in responsiveness to simple tests of acute pain (brief exposure to noxious heat). Nor did they develop any signs of morphine dependence or withdrawal when, having been treated repeatedly with large doses of morphine, they were challenged with the antagonist drug naloxone. In a behav-

ioural test known as 'place preference', which is used as an indicator of the rewarding properties of a drug, the MOR-knockout animals showed none of the normal positive responses to morphine (lowering of body temperature and reduction of inflammation, for example).

So it seems that the key effects of morphine as both an analgesic and an addictive euphoriant agent are completely lacking in the MOR-knockout mice. This is a surprising finding, as it had been assumed that the δ - and κ -opioid receptors would be involved in mediating at least the analgesic effects of morphine. However, the biological function of these receptors is now even more difficult to discern.

Some of the other findings reported by

1. Matthes, H. W. D. et al. *Nature* **383**, 819–823 (1996).
2. Hughes, J. & Kosterlitz, H. W. *Br. Med. Bull.* **39**, 1–3 (1983).
3. Pert, C. B. & Snyder, S. H. *Science* **179**, 1011–1014 (1973).
4. Hughes, J. et al. *Nature* **258**, 577–579 (1975).
5. Höllt, V. *Trends Neurosci.* **16**, 24–26 (1983).
6. Kosterlitz, H. W. & Paterson, S. J. *Proc. R. Soc. Lond. B* **210**, 113–122 (1980).
7. Keiffer, B. *Cell. Mol. Neurobiol.* **15**, 615–635 (1995).
8. Pan, Y.-X. et al. *Mol. Pharmacol.* **47**, 1180–1186 (1995).
9. Meunier, J. C. et al. *Nature* **377**, 532–534 (1995).
10. Reinscheid, R. K. et al. *Science* **270**, 792–794 (1995).

Jon Mitchell/Panos Pictures

Matthes *et al.*¹ are also remarkable. The knockout mice seemed to be unperturbed by the absence of the MOR: there was no increase in mortality, either before or after birth, and the animals showed normal circadian patterns of activity. Of particular interest is the observation that they were no more sensitive than wild-type mice to painful stimuli. The MOR-knockout mice also showed normal expression of the other two opioid receptors, and there was no change in the expression of opioid peptides in the brain.

All of these findings lend support to what has long been suspected — that the opioid system is not active under normal resting conditions. It is possible that other neuropeptide systems will also be found to show this property: they may represent a group of highly specialized

chemical signalling systems that are called into action only under particular environmental circumstances or at certain periods during the development of an animal.

The field of opioid research continues to flourish. Another new member of the opioid receptor family (opioid receptor-like, ORL₁)⁸ and another opioid peptide (orphanin FQ (ref. 9) or nociceptin¹⁰) have been discovered within the past year. The new findings¹ are an important advance and provide yet another demonstration of the power of modern genetic engineering in attacking age-old biological problems. □

Leslie L. Iversen is in the Department of Pharmacology, University of Oxford, Mansfield Road, Oxford OX1 3QT, UK.

MICROSCOPY

Random mask brightens image

Ted Dixon

CONFOCAL scanning laser microscopes have a great advantage over ordinary optical microscopes: they reject light that does not come from the focal plane. That enables them to perform optical tomography, slicing up an object into planes so that its full three-dimensional structure can be built up. But real-time imaging has always been difficult, requiring very fast beam scanners (and fast de-

light source, light expanding from the pinhole passes through a beam splitter and is focused by a microscope objective lens to a spot on a specimen (Fig. 1). Light reflected from this spot is partly reflected by the beam splitter towards a pinhole in front of a detector. Light reflected from any other position in the specimen, including parts above or below the focal plane, runs into the edge of the detector pinhole and is rejected. The objective lens forms an image of the detector pinhole and the illuminating pinhole at the same spot in the focal plane — they are said to be confocal with each other.

In order to form an image, either the specimen is moved in a raster scan under a stationary focused beam (a 'scanning stage' microscope), or the beam is scanned using mechanical mirrors so that the focus spot moves across the specimen (a 'scanning beam' microscope). Scanning-stage microscopes are relatively slow, and mechanical vibration is a problem, but scanning-beam instruments are available that capture images in real time. Because of the very short time that each point is illuminated, high beam intensity is required, and lasers are used. But that makes white-light imaging very difficult, unless red, green and blue lasers are used — a solution that considerably increases both complexity and cost.

Another way to scan the spot across the specimen is to move the illuminating pinhole. Then by synchronizing the detector pinhole with it one obtains a scanning confocal microscope. If a number of pinholes are used, the instrument acts like several scanning microscopes working in parallel, and real-time imaging can be achieved.

The first such instrument was the tandem scanning microscope designed by Petráň *et al.*² in 1968. This used the disk originally invented by Paul Nipkow³ in 1884 for encoding and decoding images for transmission over telegraph cables. In the tandem microscope, several hundred pinholes are illuminated on one side of the centre of the disk, and the light reflected from the specimen then passes through a complementary pattern of pinholes on the other side.

In effect, this is several hundred microscopes all operating in parallel. When an observer looks through the eyepiece, hundreds of focused spots, all originating from the specimen, move across the retina, and the brain averages these stimuli to produce a real-time confocal image. The microscope works well on highly-reflecting specimens, but only about 1% of the light from the source passes through the Nipkow disk, so weakly reflecting and fluorescent specimens look dim. In addition, the complicated beam path makes it difficult to align the instrument so that pinholes on one side of the disk remain truly confocal with pinholes on the other side.

One way to solve the alignment problem is to use the same pinhole for both illumination and detection (Fig. 2). The microscope objective lens focuses light from each pinhole in the Nipkow disk to a spot on the specimen, and the reflected light is collected by the same lens, and focused on the originating pinhole.

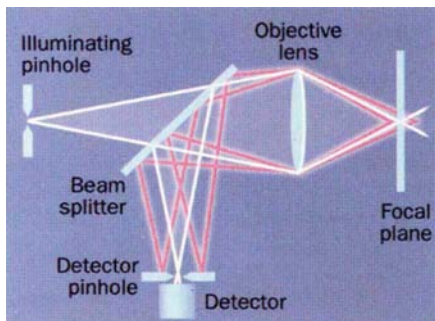


FIG. 1 A simple confocal microscope (see text).

tectors) or the use of a spinning disk containing hundreds or thousands of pinholes. Spinning disk microscopes have the advantage of being able to use white light sources, which produce real-time colour images in reflected light, but because light transmission through the disk is usually less than 1%, weakly fluorescent specimens are difficult to image. An innovative real-time microscope described by Juškaitis *et al.* on page 804 of this issue¹ changes all that.

Although commercial instruments are becoming more and more complicated, the basic principles of confocal microscopy are easy to understand. When a pinhole is illuminated by a laser or other

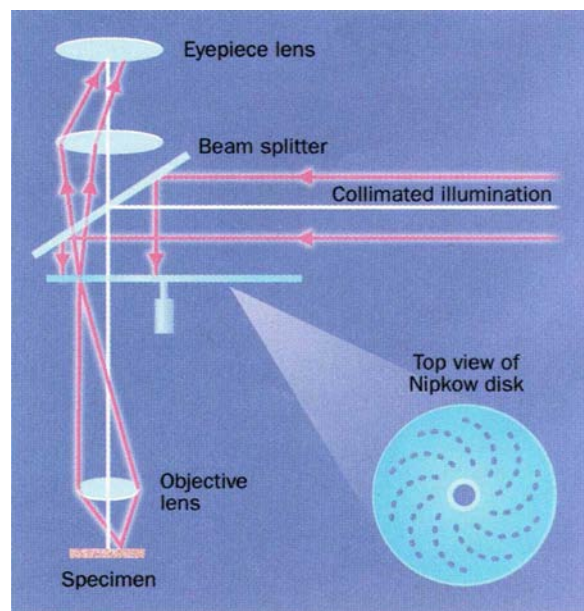


FIG. 2 A real-time confocal microscope (see text).

Although easier to align than the tandem system, this instrument has a different disadvantage. Light reflected from the top of the disk enters the eye as well as the light coming back from the specimen, reducing contrast in the image. Most of this stray light can be removed by placing a polarizer in the incoming beam from the arc lamp, a quarter-wave plate below the Nipkow disk, and a second polarizer above the beam splitter. The second polarizer can then be rotated to block light from the top of the disk but transmit light returning from the specimen, which has passed through the quarter-wave plate twice, rotating its polarization by 90 degrees. Unfortunately, this also reduces the intensity of the signal.

Another improvement was made by Xiao *et al.*⁴. They covered the Nipkow disk with black chrome (with a reflectivity of only a few per cent), and tilted it so that most of the light reflected from it passed into a light trap instead of into the optics of the instrument.

But even with all of these improvements, the real-time microscope suffered from a poor light budget. Pinhole size was constrained by the necessity of using pinholes that were small enough for the light diffracted from each pinhole to just fill the entrance pupil of the objective lens, and small enough to provide true confocal imaging, while at the same time letting as much light through as possible. Pinholes couldn't be less than about ten radii apart, to eliminate cross-talk between adjacent optical paths.

This seemingly intractable problem has been solved by Juškaitis *et al.*¹. They describe an aperture mask in which the array of holes is packed closely together, allowing up to 50% of the available light to be used, but coded in random sequences, so that the openings and closings of adjacent apertures are completely uncorrelated with one another. The result, after some processing, is a confocal image superimposed on a conventional image (produced by all the cross-talk between the uncorrelated aperture sequences). Another conventional image is obtained separately, and subtracted electronically from the combined image to yield a real-time confocal image. This major advance in confocal microscopy will be particularly important whenever weakly fluorescent or weakly reflecting specimens must be imaged in real time. □

Ted Dixon is in the Department of Physics, University of Waterloo, Ontario, Canada N2L 3G1.

Molecular sexing of birds

Kate Lessells and Christa Mateman

THE flamboyant plumage of adult male pheasants, mallards and bullfinches reveals the sex of the wearer to even the most casual observer, but separating bird sexes is not always so easy — there are no size or plumage differences between adult males and females in many species, and newly hatched chicks can virtually never be sexed in this way. As a result, ornithologists are constantly on the lookout for an 'off-the-shelf' technique with which to sex their study species. A gene on the W chromosome of birds, initially found by Richard Griffiths and colleagues^{1,2}, and now investigated by his group³ and by Hans Ellegren⁴, comes the closest yet to providing a universal method for sexing birds.

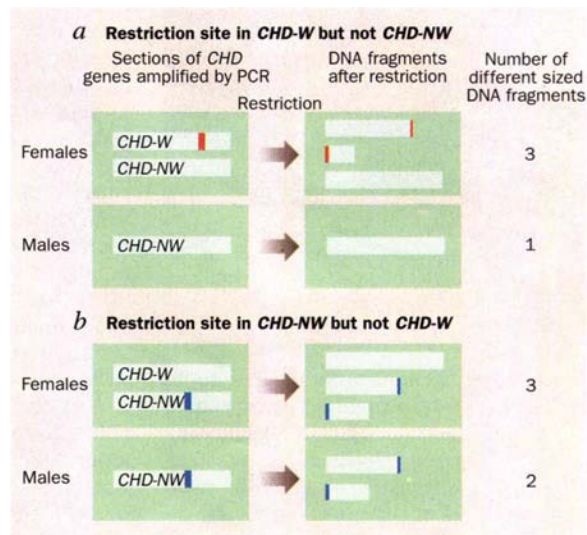
At the chromosomal level, sex determination in birds is similar to that in mammals (including humans), except that it is the males who have a pair of identical sex chromosomes (ZZ), and the females who have one Z and one W chromosome (ZW). Consequently, all DNA sequences found in males also occur in females, but DNA sequences on the W chromosome may be unique to females. Almost all molecular-biological methods for sexing birds exploit such female-specific sequences, but to date, each of the sequences that has been used is restricted to one or a few taxonomically close species.

As a result, sexing a new species involves a lengthy trial-and-error process to identify a suitable sequence. In contrast, the avian *CHD-W* (chromodomain-helicase-DNA-binding W-linked) gene identified by Griffiths *et al.*³ apparently occurs on the W chromosome of all birds, except ostriches and their close allies, and hence is a (near-) universal marker for sex in birds.

To use the *CHD-W* gene for sexing birds, a laboratory technique is needed that reveals the presence or absence of the gene. But a second version of the *CHD* gene — *CHD-NW* — is located on one of the other chromosomes, and so occurs in both males and females. Because *CHD-NW* differs from *CHD-W* at only a small proportion of base pairs, any sexing technique must be able to discriminate the

slight differences between the two genes. However, the existence of *CHD-NW* also provides an invaluable internal control to check that *CHD-W* is absent because the bird is male, and not simply because of technical failure.

Various techniques are available^{3,4}, but those based on the polymerase chain reaction (PCR) are generally preferred. PCR is fast, and because it amplifies the DNA,



How a restriction enzyme can be used to distinguish between males and females. Female birds have both the *CHD-W* and *CHD-NW* genes, which differ in sequence at a small proportion of base pairs, whereas males have only the *CHD-NW* gene. A restriction enzyme cuts DNA at a specific short sequence of base pairs. *a*, If a restriction enzyme is chosen which cuts at a particular target sequence (shown in red) that occurs in *CHD-W* but not in *CHD-NW*, restriction will result in fragments of three different sizes in females, and of only a single size in males. *b*, If the target sequence (shown in blue) occurs in *CHD-NW* but not *CHD-W*, restriction will generate fragments of three different sizes in females and of two sizes in males. The number of different-sized fragments, and hence the sex of an individual, can then be determined by gel electrophoresis.

only tiny initial amounts of DNA are needed — enough DNA can be extracted from feathers², or from blood samples so small that even newly hatched chicks can be sexed.

Further choice of technique largely depends on the expertise of the researcher. Those who aren't molecular biologists need a method that makes minimal technical demands and will work in their species with the least tweaking. For them, one possibility is to amplify a section of both of the *CHD* genes (when present) using a pair of PCR primers developed by Griffiths², cut the amplified DNA with a restriction enzyme (see figure), and separate the resultant DNA fragments according to size using gel electrophoresis³. The part of the technique that may need to be modified for each new species is the

1. Juškaitis, R., Wilson, T., Neil, M. A. A. & Kozubek, M. *Nature* **383**, 804–806 (1996).
2. Petran, M., Hadravsky, M., Egger, D. & Galambos, R. *J. Opt. Soc. Am.* **58**, 661–664 (1968).
3. Nipkow, P. German patent no. 30,105 (1884).
4. Xiao, G. Q., Corle, T. R. & Kino, G. S. *Appl. Phys. Lett.* **53**, 716–718 (1988).

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Many bird species, such as the rainbow lorikeet (*Trichoglossus haematodus*), cannot be sexed by looks alone. Now, a new technique developed by Griffiths *et al.*³ provides a molecular solution to this problem.

choice of restriction enzyme. This could be done either by a trial-and-error search, or by the more technically demanding DNA sequencing. However, the restriction enzyme suggested by Griffiths *et al.* works in all 14 of the phylogenetically diverse species tested³. The main drawbacks are that there is no way to check that the restriction enzyme has cut the DNA, and that the PCR primers can also amplify human DNA, making contamination an ever-present risk.

Those with the expertise and equipment to sequence the *CHD* genes can use

other techniques that simultaneously avoid these problems and almost guarantee the eventual ability to sex a particular species. The most attractive option is to use the sequence information to design PCR primers that amplify fragments of appreciably different size from the *CHD-W* and *CHD-NW* genes⁴. This circumvents the need for restriction digestion, and means that sex can be determined using PCR followed by electrophoresis of the amplified DNA.

The function of the *CHD* genes (global regulation of transcriptional activation at the chromatin level) is largely irrelevant to their use in sexing birds, although the genes may, of course, be involved in the physiological determination of sex. The fact that the DNA sequence does encode a protein is, however, the reason that the genes occur in almost all birds. Much of the rest of the W chromosome apparently consists of repetitive 'junk' DNA. Freed from the selective constraints imposed by coding for a protein, such DNA can evolve rapidly and diverge between species. It is likely that the species-specificity of probes and primers previously used for molecular sexing results from the detection of such non-coding DNA. In contrast, the *CHD* genes are remarkably conserved — although mammals and birds diverged more than 200 million years ago, a mouse gene that is homologous to the chicken *CHD-W* gene is 82.9 per cent identical at the nucleotide level, and 95.6 per cent identical at the amino-acid level⁴.

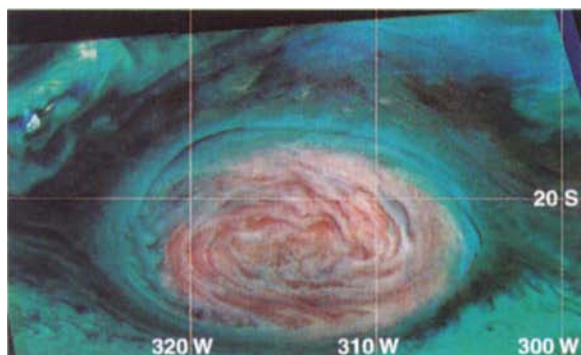
The availability of increasingly simple and generally applicable techniques for sexing birds — such as those based on the *CHD* genes — is having a great impact on avian research. In conservation, the ability to sex individuals from such unpromising

material as a single moulted feather enables individuals of the appropriate sex to be chosen for release into the wild², captive breeding or movement to a new area. Survival, reproductive rates and dispersal often differ between the sexes, so the ability to sex individuals is essential to understanding demography. In evolution, sex ratios may be adaptively modified within species in relation to factors such as the timing of breeding⁵, cooperative breeding⁶ and parental quality⁷. Techniques based on PCR mean that large numbers of birds can be sexed early in life, allowing differential survival to be distinguished from variation in the primary sex ratio. Looking to the future, the new techniques can potentially be used at, or soon after, the meiosis that produces an ovum — this meiosis determines whether the ovum receives a Z or a W chromosome, and hence the sex of the resultant embryo. Investigating such fundamental questions as the mechanism that allows females to control the sex of their offspring is therefore brought within reach by the development of molecular techniques for sexing birds. □

Kate Lessells and Christa Mateman are at the Netherlands Institute of Ecology, Boterhoeksestraat 22, PO Box 40, 6666 ZG Heteren, The Netherlands.

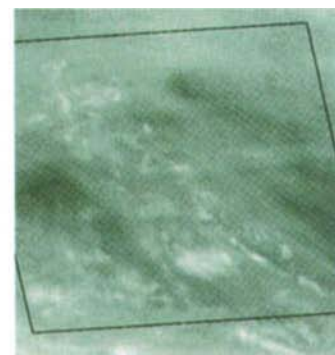
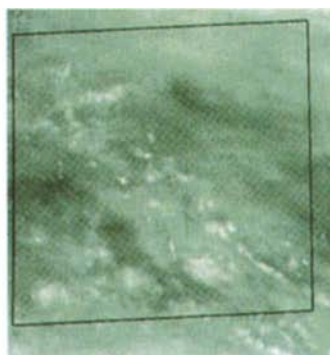
1. Griffiths, R. & Tiwari, B. *Proc. Natl Acad. Sci. USA* **90**, 8324–8326 (1993).
2. Griffiths, R. & Tiwari, B. *Nature* **375**, 454 (1995).
3. Griffiths, R., Daan, S. & Dijkstra, C. *Proc. R. Soc. Lond. B* **263**, 1251–1256 (1996).
4. Ellegren, H. *Proc. R. Soc. Lond. B* (in the press).
5. Daan, S., Dijkstra, C. & Weissing, F. J. *Behav. Ecol.* (in the press).
6. Komdeur, J. *Proc. R. Soc. Lond. B* **263**, 661–666 (1996).
7. Ellegren, H., Gustafsson, L. & Sheldon, B. C. *Proc. Natl Acad. Sci. USA* **93**, 11723–11728 (1996).

Changeable, with a chance of showers



JUPITER might have his thunderbolts after all. More importantly, the water that planetary scientists expected to find there, but that the Galileo probe denied them, might be hiding in thunderclouds. According to a theory presented last week at the AAS meeting in Tucson, Arizona by Andrew Ingersoll and colleagues, the probe just happened to fall through a dried-out downdraft.

The two pictures on the right were taken 70 minutes apart



by the probe's mothership, the Galileo spacecraft. The area is 6,000 km on a side, to the north and east of the Great Red Spot (pictures from M. J. S. Belton *et al. Science* **274**, 377–385; 1996). The clouds clearly change rapidly, like earthly thunderstorms, and they are very tall (30 km or so). It is difficult to produce such powerful convection systems using anything but water.

Stephen Battersby

Reprogramming transcription

David W. Mount

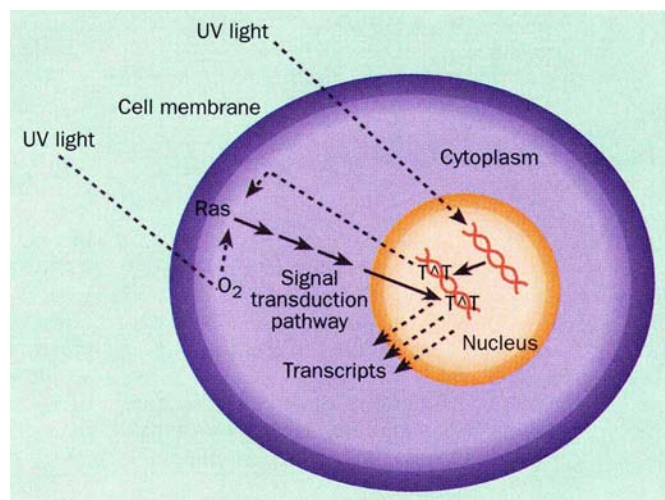
EXPOSURE of living organisms to the ultra-violet (UV) component of sunlight leads to biological damage. Wavelengths of UV light that reach the surface of the Earth (UV-B) and shorter, artificially generated wavelengths (UV-C) produce lesions in DNA — namely cyclobutyl pyrimidine dimers and 6-4-pyrimidinone dimers — which block DNA replication and transcription. A variety of mechanisms for repairing or circumventing the damage are present in both prokaryotic and eukaryotic organisms, and the genes that regulate these mechanisms are often induced by DNA damage¹. It has also been proposed that UV light generates oxidation products near the cell membrane in eukaryotic cells. These oxidation products activate a signal-transduction pathway that leads to the increased expression of specific genes²⁻⁴. Now, on page 826 of this issue, Conconi *et al.*⁵ propose that UV light induces the expression of several defence genes in tomatoes through such a pathway.

In *Escherichia coli* a system exists for the induction of specific genes by UV light. Known as the SOS response, this system involves the processing of a damaged DNA molecule to create a signal for the rapid destruction of the LexA repressor by proteolytic cleavage. LexA cleavage leads to the derepression of many genes involved in DNA repair, recombination, mutagenesis and the inhibition of cell division, which may then act in a concerted manner to remove the damage or to minimize its harmful effects¹.

In the yeast *Saccharomyces cerevisiae*, DNA damage leads to increased transcription of a number of diverse DNA repair, replication and recombination

genes⁶. Transcription is regulated by cell-cycle checkpoint genes⁷, some of which process the DNA damage⁸, and by Dun1 protein kinase⁹. These results indicate that a series of protein-phosphorylation events is involved in the response of yeast cells to DNA damage.

In animal cells, UV light also induces a characteristic UV response^{2,10}. A signal cascade involving membrane-associated



Activation of signal-transduction pathways by UV light. In one model, UV light produces DNA damage in the nucleus. The damaged nucleus then sends an activating signal to a signal-transduction pathway mediated by Ras, which is located in the cytoplasm near the cell membrane. A different model proposes that UV light produces free radicals near the cell membrane, causing oxidative stress. This activates the same pathway, which transmits a signal through the cytoplasm back to the nucleus, eventually activating transcription factors and inducing new transcripts. The affected signalling pathway also responds to other types of signals, such as hormones and growth factors.

Ras protein followed by cytoplasmic protein kinases activates the transcription factors AP1 and NF- κ B to increase transcription of their target genes. A similar UV response occurs in yeast: membrane-associated Ras activates a signalling pathway that leads to increased translation and enzyme activity of GCN4. GCN4 is the yeast counterpart of AP1, and it stimulates increased expression of the amino-acid-synthesizing genes *HIS3* and *HIS4* (ref. 4). Now the induction of tomato defence genes described by Conconi *et al.*⁵ seems to be a similar type of UV response in plants. However, the nature of the inducing mechanism and the exact function of these responses are not clear.

Two signalling pathways for the UV response can be envisaged (see figure). In the first, DNA damage generates the primary signal¹⁰. A secondary signal is then transmitted to the cytoplasm, thereby activating specific signal-transduction pathways. In the second signalling pathway,

UV light induces oxidative stress and the production of free radicals near the cell membrane. These molecules elicit changes in membrane-associated proteins or lipids, thus activating the same signal-transduction pathways²⁻⁴, but by a different mechanism. In both pathways, the signalling component activates transcription factors, leading to the increased transcription of specific genes.

The second pathway has gained support from the observation that NF- κ B can be activated in enucleated cells which do not produce a nuclear DNA-damage signal³. Moreover, the activation of NF- κ B is reduced by growing normal cells in a medium containing *N*-acetylcysteine, which is an antioxidant². Induction of the pathogen-responsive gene *PRI* in tobacco plants is also reduced by treatment with antioxidants, but not by treatments that remove UV-induced DNA damage by a photoreactivation mechanism¹¹. Although these studies suggest that DNA damage is not involved in the induction of signalling pathways, they do not eliminate the possibility of a role for residual damage to mitochondrial or nuclear DNA. Conconi *et al.*⁵ show that at doses of UV radiation that are sufficient to induce plant defence genes, an average of 0.2 cyclobutyl pyrimidine dimers are introduced per gene. Such a high level of DNA damage is probably enough to produce marked biological effects.

The complex changes in gene expression and cell-cycle regulation that are caused by DNA damage are almost certain to be important in these UV-response mechanisms. In addition, the high, acute doses of UV-C light that are often used in these experiments may not provide biologically relevant information about the cellular response mechanisms — they may not be active at these doses. Do these UV responses represent the activities of a dying cell or a useful biological function such as a repair mechanism?

HeLa cells can be made considerably more sensitive to UV light by blocking the signalling pathway with tyrosine kinase inhibitors². And in yeast, mutant strains in which *GCN4* is constitutively active are more resistant to UV light than are wild-type strains, whereas mutant strains in which *GCN4* cannot be activated are more sensitive⁴. So the UV response has a protective effect, but it is not clear what this effect might be.

Exposure of cells to UV light might increase the levels of a protective mol-

1. Friedberg, E. C., Walker, G. C. & Siede, W. *DNA Repair and Mutagenesis* (ASM, Washington DC, 1995).
2. Devary, Y., Gottlieb, R. A., Smeal, T. & Karin, M. *Cell* **71**, 1081-1091 (1992).
3. Devary, D., Rosette, C., DiDonato, J. A. & Karin, M. *Science* **261**, 1442-1445 (1993).
4. Engelberg, D., Klein, C., Martinetto, H., Struhl, K. & Karin, M. *Cell* **71**, 381-390 (1994).
5. Conconi, A., Smerdon, M. J., Howe, G. A. & Ryan, C. A. *Nature* **383**, 826-829 (1996).
6. Aboussekhra, A. *et al.* *EMBO J.* **15**, 3912-3922 (1996).
7. Kiser, G. L. & Weinert, T. A. *Mol. Biol. Cell* **7**, 703-718 (1996).
8. Lydall, D. & Weinert, T. *Science* **270**, 1488-1491 (1995).
9. Zhou, Z. & Elledge, S. V. *Cell* **75**, 1119-1127 (1993).
10. Herrlich, P. *et al.* *Rev. Physiol. Biochem. Pharmacol.* **119**, 187-223 (1992).
11. Green, R. & Fluhr, R. *Plant Cell* **7**, 203-212 (1995).
12. Chappell, J. & Hahlbrock, K. *Nature* **311**, 76-78 (1984).

ecule that prevents damage to DNA or cytoplasmic constituents. In plants, for example, UV light induces the transcription of genes that are involved in synthesis of UV-absorbing molecules such as flavonoids and phenylpropanoids¹². Perhaps there is some evolutionary advantage to the UV-mediated induction of the same plant defence genes as those that are normally induced by insect and pathogen attack. A possible example would be if the intensity of UV-B were correlated with insect activity. Also, activation of signal-transduction pathways that regulate cell growth and division might have a beneficial effect on the survival of UV-damaged cells.

Finally, DNA repair is targeted to the transcribed strand of the genes that are undergoing transcription¹, so the induction of biologically important genes by the UV-response pathway will increase the speed at which lesions are excised from these genes. This should also have the effect of reducing the overall level of genome damage. Thus, one function of the UV response in eukaryotic cells could be to increase the rate of DNA repair through an increase in the transcriptional programme of the cell. □

David W. Mount is in the Department of Molecular and Cellular Biology, University of Arizona, Tucson, Arizona 85721, USA.

SUPERFLUIDS

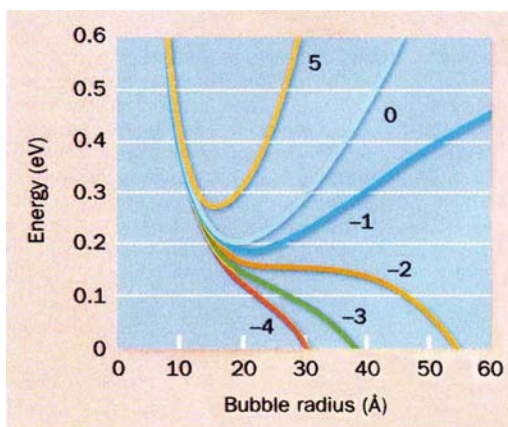
Exploding electron bubbles

Peter McClintock

ELECTRON bubbles, used in laboratories throughout the world for probing the unusual properties of liquid helium, can be made to explode by the application of negative pressure, according to investigations by Classen *et al.* published last month¹.

Helium is unique among liquids in that it can be cooled to arbitrarily low temperatures without solidifying—in principle, right down to zero kelvin. As it cools, the average speed of its atoms in random thermal motion decreases, and their de Broglie wavelengths (inversely proportional to momentum) correspondingly increase. When the waves of neighbouring atoms strongly overlap, their motion becomes correlated, and the liquid undergoes Bose-Einstein condensation² to a superfluid state with a whole set of exotic properties that have already kept physicists busy for more than half a century. For example, the superfluid flows without friction, exerts no drag on a moving object, tends to climb out of an open-topped container, and can only rotate at certain quantized angular velocities relative to the fixed stars.

This year's Nobel prize in physics went to David Lee, Douglas Osheroff and Robert Richardson for their 1972 discovery of superfluidity in ³He, since when it has been intensively studied using nuclear magnetic resonance. Its isotopic sibling ⁴He cannot be investigated in this way because its nucleus is non-magnetic.



Calculated energy of an electron bubble in liquid helium as a function of its radius, for pressures between 5 and -4 atmospheres¹. No equilibrium bubble radius exists for pressures below -2 atmospheres, so they explode.

For this reason, electron bubbles—often misleadingly known as negative ions—have been widely used over the years to probe its superfluidity.

Their main features are well established³, and can be understood in terms of the quantum-mechanical requirement that the lowest permitted kinetic energy of a particle depends on the volume it is confined in: make the volume smaller, and the so-called zero-point energy rises. This decidedly non-classical effect gets bigger as the mass of the particle decreases, and is especially pro-

nounced for a very light particle such as an electron. Thus, an excess electron in liquid helium can decrease its energy by carving out a small spherical cavity for itself.

Enlargement of the bubble reduces the zero-point energy, but also involves doing work against surface tension and the external pressure. So there is a compromise (equilibrium) radius at which the total bubble energy is minimized. At zero pressure, this is about 19 Å, so rather than having a naked electron in liquid helium, one finds a mesoscopic, charged, spherical bubble of vacuum with an effective hydrodynamic mass equal to half the mass of displaced fluid (a few hundred helium atomic masses).

Electron bubbles of this kind have been used to measure the quantum of circulation⁴, that is, the liquid's allowed increments of rotational motion, which take physical form as quantized vortex lines. They have also provided techniques for imaging the vortex lines photographically⁵, for measuring the Landau critical velocity at which superfluidity breaks down⁶, and for demonstrating that vortex creation occurs through a form of macroscopic quantum tunnelling⁷ closely analogous to processes recently observed in magnetic systems (see ref. 8 for a brief review).

What Classen *et al.* have now done is to take this familiar probe, subject it to negative pressure, and see what happens. Their experimental technique involves the use of focused ultrasound, which causes the pressure on the bubble to rise and fall cyclically above and below ambient pressure. For large enough amplitudes, the pressure minima can become negative.

What should happen to an electron bubble near the focus of the ultrasound, where the pressure swings are largest? The figure plots the calculated bubble energy as a function of radius for several pressures¹. Above a critical pressure of about -2 atmospheres, a minimum exists in the energy at some radius, determining the equilibrium state of the bubble. For lower pressures, however, there is evidently no minimum, and one can conclude that there is then no stable configuration for the bubble. So if the pressure minima induced by the ultrasound go below -2 atmospheres, the bubble will start to expand without limit; in other words, to explode. In this set-up, disaster is prevented by the high-pressure part of the cycle, which recompresses the bubble and restores its stability.

Erratum

In the article 'Tangle disentanglement' in the 10 October issue (383, 476-477; 1996), ref. 1 was cited as from *Mol. Psychol.* The correct title of the journal is *Molecular Psychiatry*.

- Classen, J., Su, C.-K. & Maris, H. J. *Phys. Rev. Lett.* **77**, 2006-2008 (1996).
- London, F. *Nature* **141**, 643-644 (1938).
- Fetter, A. L. in *The Physics of Liquid and Solid Helium* (eds Bennemann, K. H. & Ketterson, J. B.) 207-305 (Wiley, New York, 1976).
- Rayfield, G. W. & Reif, F. *Phys. Rev. A* **136**, 1194-1208 (1964).
- Yarmchuk, E. J., Gordon, M. J. V. & Packard, R. E. *Phys. Rev. Lett.* **43**, 214-217 (1979).
- Ellis, T. & McClintock, P. V. E. *Phil. Trans. R. Soc. Lond. A* **315**, 259-300 (1985).
- Hendry, P. C. *et al. Phys. Rev. Lett.* **60**, 604-607 (1988).
- Stamp, P. C. E. *Nature* **383**, 125-126 (1996).

Classen and colleagues used a β -emitting radioactive source to inject electrons into the liquid helium, and looked for exploding bubbles by shining a laser through the liquid. Electron bubbles are normally far too small to scatter light, but in their vastly expanded exploded state they can be expected to scatter it effectively. In the event, the authors were able not only to confirm the existence of the exploding bubbles, but also to determine the probability of explosions as a function of the ultrasound amplitude, and show that it fits expectations.

It would not surprise me if exploding electron bubbles came in useful for

something — perhaps non-destructive imaging of an electron cloud, to watch where they go in the helium? But essentially this is good basic science. Although there are some unexpected features of the results that need further investigation — notably phenomena observed below 1 K, possibly caused by bubbles getting stuck on quantized vortices — the work demonstrates convincingly that the normally placid electron bubble can sometimes explode. □

Peter McClintock is in the School of Physics & Chemistry, Lancaster University, Lancaster LA1 4YB, UK.

DEVELOPMENTAL BIOLOGY

High hops of transgenic frogs

J. M. W. Slack

THOSE of us who work in developmental biology, and use the frog *Xenopus laevis* as our experimental organism, are used to being stopped in the corridor and told that *Xenopus* are useless because "They have no genetics". Other model organisms such as *Drosophila*, mouse, zebrafish and *Caenorhabditis elegans* have been chosen mainly because they are suitable for genetic experiments, whereas frogs are used because their embryos are large and appropriate for micromanipulation. *Xenopus* do, of course, have genes. But they are not suitable for mutagenesis or breeding experiments because of their long generation time, and because they have two similar copies of each chromosome, having apparently undergone a genome duplication about 30 million years ago.

Now, however, one of the most important techniques available to those who work with mice or *Drosophila* has also become applicable to frogs. This is transgenesis, or the introduction into the organism of genes engineered in the laboratory. The new methods are described in *Development* by Kroll and Amaya¹; and given the high level of conservation of developmental mechanisms between different animal groups, their application should tell us more not only about frogs but about animal development generally.

Transgenic mice are made routinely by injecting DNA into a pronucleus of the fertilized egg, and similar experiments have been performed with frogs for many years². The trouble is that only a minority of cells in the embryo express the transgene. The reasons for this 'mosaic' expression are not fully understood, but it effectively limits DNA-based experiments to three types — use of reporter genes to study the tissue specificity of modified promoters³, or to probe the intracellular regulatory environment in different re-

gions of the embryo⁴; or overexpression of secreted factors that can affect not only the cells in which the gene is active but also surrounding tissues⁵.

It has not been possible to produce consistent changes in the anatomy of the embryo by the overexpression of genes whose products are retained within the cell, for example receptors, kinases or transcription factors, because too few of the cells in the body are transformed. Many groups have used RNA-based overexpression which does not suffer from this problem, but injected messenger RNA is translated from the time of injection and this effectively limits studies to the earliest stages of development because later

events are often obscured by interference from earlier ones.

Another potential route to transgenesis is nuclear transplantation. This was first successfully performed in 1952 by Briggs and King⁶, who showed that blastula nuclei could be transplanted into enucleated eggs of another frog species, *Rana pipiens*, and could support development up to adulthood. Subsequent work, mainly using *Xenopus*, showed that, although nuclei from progressively later stages supported development less well than did blastula nuclei, it was possible to obtain a few well-differentiated tadpoles even using nuclei from fully differentiated cells grown in primary culture⁷. Building on this, transgenes have been introduced into *Xenopus* cell lines and the nuclei used for the transplantations^{8,9}. But although the transgenes could be expressed throughout the embryo, the cell nuclei could not support normal development. This is probably because *Xenopus* tissue culture is not a highly developed technology and all existing cell lines have abnormal karyotypes. When the host nucleus was left in place, survival was better but expression of the transgene then tended to be mosaic.

Kroll and Amaya¹ have circumvented the problem of aneuploidy by the simple expedient of introducing the transgene into sperm instead of tissue culture cells (Fig. 1). Using green fluorescent protein as a marker, the authors show that their embryos express the transgene in all cells while the same gene on injected DNA is expressed only in a minority of cells. When tadpoles were reared to 1–2 months of age, the transgenes were still

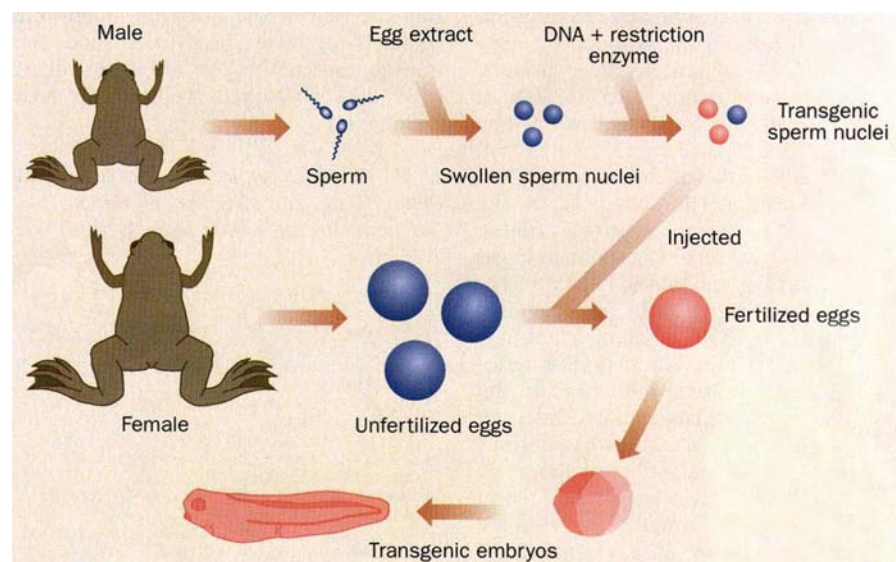


FIG. 1 The new procedure for making transgenic *Xenopus*, developed by Kroll and Amaya¹. Sperm nuclei are treated with an egg extract to swell them and decondense the chromatin. The transgene is then added along with a restriction enzyme to make complementary breaks in the sperm DNA. The sperm nuclei are injected into unfertilized eggs, obtained from females by injection of chorionic gonadotrophin. Although eggs can be activated to begin development merely by pricking, if no sperm, or more than one sperm, has been injected, then the cleavages are aberrant and development soon ceases.

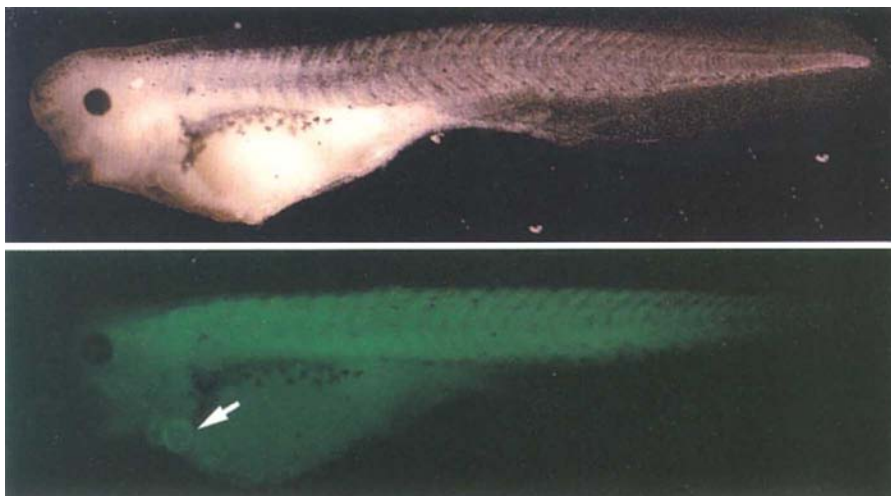


FIG. 2 A transgenic tadpole produced by Kroll and Amaya¹. It contains a gene for green fluorescent protein driven by a muscle-specific promoter. Green fluorescence is apparent both in the myotomes and in the heart (arrow). (Reproduced from ref. 1.)

present in all cells and were still expressed. The transgenes appear to be present in 5–35 copies per cell, as single copies or as short concatamers at 4–8 sites in the genome. By using muscle- and neuron-specific promoters, the authors also show that the expression of the transgene can be directed to the appropriate tissues (Fig. 2).

The main importance of transgenesis is that it enables much more sophisticated overexpression experiments to be carried out on embryos. The huge range of tissue-specific promoters and inducible elements available means in principle that any gene can now be turned on or off at any stage of development or at any position within the embryo. Although it will still not be possible to create mutants of endogenous genes as in mice, it is possible to inhibit a very wide range of gene products by overexpression of dominant negative variants, so workers should now also be able to inhibit the activity of most genes at any desired time and place.

Kroll and Amaya¹ have themselves started to explore the potential of the technique. During the gastrula stages, they have overexpressed a dominant negative receptor that inhibits all known receptors for fibroblast growth factor (FGF). They show that FGF signalling is not only required for the initial induction of mesoderm in the blastula, but also for the subsequent maintenance of mesodermal genes. This confirms the conclusions drawn from more indirect experiments^{5,10}. They have also demonstrated that neural induction occurs normally; so although FGFs have been shown to act as neural inducers *in vitro*^{11,12}, it seems unlikely that FGF signalling is necessary for this process in normal development.

A further upshot of the new approach is that some permanent lines of transgenic frogs will probably be created. An obvious requirement is a line with marked cells,

for example one expressing β -galactosidase or green fluorescent protein in all cells. Embryos from such frogs could serve as a source of labelled cells for grafting experiments without the need to inject donors with fluorescent lineage labels. Another likely application is the incorporation of temperature-sensitive oncogenes which could allow the immortalization of primary cultures from late-embryo tissues. The scope for this should be wide — frogs are poikilothermic, and so both the animals and the cell lines can be kept over a much wider range of temperatures than is possible for mammals.

Every generation of developmental biologists thinks that frogs have had their day and done their bit, and are just about to be superseded by some other organism. But the old frogs carry on being the vehicles for new discoveries, and the transgenesis method should ensure their survival as lab animals well into the next millennium. □

J. M. W. Slack is in the Developmental Biology Programme, School of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK.

1. Kroll, K. L. & Amaya, E. *Development* **122**, 3173–3183 (1996).
2. Gurdon, J. B. & Melton, D. A. *Annu. Rev. Genet.* **15**, 189–218 (1981).
3. Mohun, T. J., Garrett, N. & Gurdon, J. B. *EMBO J.* **5**, 3185–3193 (1986).
4. Watabe, T. et al. *Genes Dev.* **9**, 3038–3050 (1995).
5. Isaacs, H. V., Pownall, M. E. & Slack, J. M. W. *EMBO J.* **13**, 4469–4481 (1994).
6. Briggs, R. & King, T. J. *Proc. Natl Acad. Sci. USA* **38**, 455–463 (1952).
7. Gurdon, J. B., Laskey, R. A. & Reeves, D. R. *J. Embryol. Exp. Morphol.* **34**, 93–112 (1975).
8. Kroll, K. L. & Gerhart, J. C. *Science* **266**, 650–653 (1994).
9. Chan, A. P. & Gurdon, J. B. *Int. J. Dev. Biol.* **40**, 441–451 (1996).
10. Amaya, E., Stein, P. A., Musci, T. J. & Kirschner, M. W. *Development* **118**, 477–487 (1993).
11. Kengaku, M. & Okamoto, H. *Development* **119**, 1067–1078 (1993).
12. Lamb, T. M. & Harland, R. M. *Development* **121**, 3627–3636 (1995).

DAEDALUS

Hydrogen for ever!

OLD age is often a slow, undignified, pitiful decline. Our primitive ancestors mostly reproduced early and died young. Evolution had no incentive, so to speak, to defend the few elderly survivors against senile decay. But what mounts the attack? One school of thought blames oxygen. Many key metabolic processes generate oxygen radicals — energetic, destructive molecular species. They gradually sabotage our DNA, sapping our powers towards ultimate collapse.

This explains why mice, fish and some insects live longer if you starve them. Their metabolism slows, they use less oxygen and generate fewer oxygen radicals. An obsessive dieter or anorexic might thus add a few more painful years to his or her allotted span — but would it be worth it? Daedalus is now attacking the problem from the oxygen side.

He noted last week that oxygen radicals can be destroyed by hydrogen. Hydrogen is quite harmless. Indeed, about two-thirds of us generate it in our intestines, by fermentation. The other one-third, like cows and most other mammals, generate methane. Daedalus suspects evolution at work. With the invention of language, old people suddenly found a role — as custodians of hard-won tribal knowledge. Natural selection began to favour those tribes whose elders stayed alive; and one of its tricks was to switch intestinal fermentation from methane to hydrogen. The hydrogen diffuses through the bowel wall into the bloodstream, and gets all round the body, reducing oxygen radicals wherever they may lurk.

This trick has evolved so recently that not all of us use it yet. So DREADCO demographers are dividing a large cohort of middle-aged folk into hydrogen and methane generators, to compare their rates of ageing. No embarrassing bowel sampling is needed; once in the bloodstream, these gases reach the lungs and can be detected on the breath. The team is also studying the medical and pension records of rocket engineers, deep-sea divers, margarine workers and others exposed to hydrogen.

If hydrogen turns out to be the elixir of long (if not eternal) life, Daedalus will try to boost its level in the body. The best long-term methods are biological. Hydrogen can be released from glucose by an enzyme reaction, which with luck might be induced to go in the body. But natural hydrogen flatulence remains the best bet. DREADCO biologists hope to modify the gut organisms responsible, so that they will colonize all of us, and go to work with enhanced zest. A long and active life, free of senile decay, would be well worth the social embarrassment of copious bowel gas.

David Jones

Insect protection against viruses

SIR — Insects exhibit both humoral and cellular immune responses that are effective against various pathogens and parasites including fungi, protozoa, bacteria, nematodes and parasitic hymenoptera¹. No insect immune response effective against viral infections has yet been described. Using a recombinant of the baculovirus *Autographa californica* M nuclear polyhedrosis virus containing the *lacZ* reporter gene (AcMNPV-hsp70/*lacZ*)², we investigated the physiological basis of resistance in *Helicoverpa zea*, an important agricultural pest and a highly refractory host³. We discovered that the larvae of this insect are actually very susceptible to infection by AcMNPV, but infected cells are encapsulated by haemocytes and subsequently cleared. These results demonstrate that the insect immune response can be effective against viral pathogens.

We orally inoculated newly moulted third-instar *H. zea* with 2,200 polyhedra of AcMNPV-hsp70/*lacZ* using a microapplicator (ISCO) and maintained the insects individually at 28±1 °C on an artificial diet. Larvae were killed at various hours post-inoculation (h.p.i.), and larval tissues were processed for elucidation of the blue *lacZ* signal and examined to determine the extent of viral infection². During the first 20

h.p.i., the proportion of *lacZ*-expressing insects increased to 96%, and early pathogenesis was nearly identical to that described for *H. virescens*, a closely related species highly susceptible to fatal AcMNPV infections⁴. Between 48 and 72 h.p.i., however, the proportions of *H. zea* that were positive for *lacZ* decreased to 40%, suggesting that some larvae had cleared their infections. In many insects killed at 30 h.p.i. and at later times, we observed small brown patches colocalized with blue *lacZ* signals in the epidermis of tracheae associated with the midgut (*a, b* in the figure). These infected tracheae are surrounded by aggregations of haemocytes, often containing capsules encompassing the host cells expressing *lacZ* (*c* in the figure). We never encountered a temporal decline in *lacZ*-positive larvae nor observed a haemocyte encapsulation response while examining thousands of dissected susceptible hosts challenged with the same viral construct^{2,4,5}. Morphologically, the capsules in *H. zea* are identical to those attributed to the insect cellular immune response in which circulating haemocytes surround, immobilize and kill invading pathogens and parasites¹.

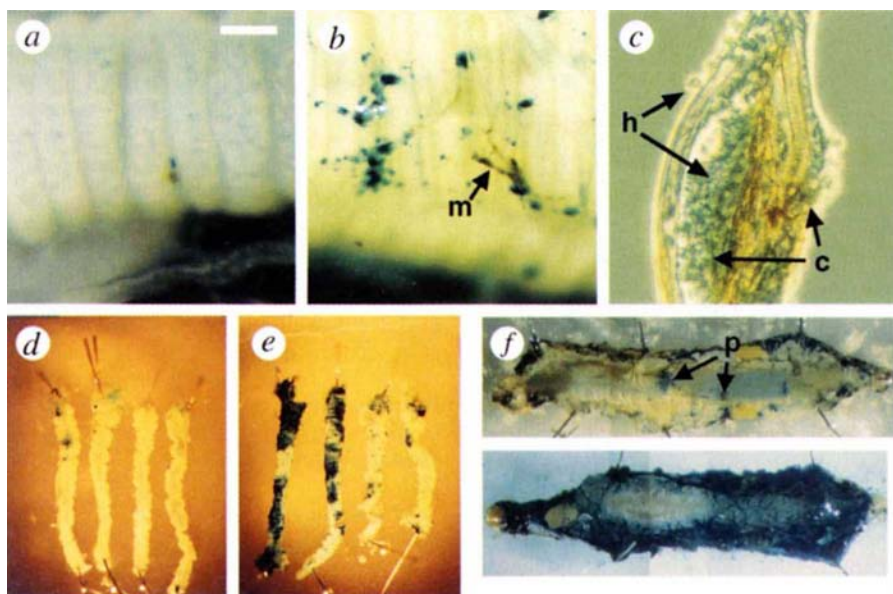
To assess the extent to which the immune response of *H. zea* is responsible

for halting the spread of AcMNPV infection, we monitored the distribution of *lacZ* expression within tissues following application of chemical and biological agents known to immunosuppress larval lepidopterans. Chemically, we compromised the immune responses of fourth-instar *H. zea* with intrahaemocoelic injections of diethylthio-carbamic acid (DDCA)⁶. In these experiments, 24 h after injection of DDCA, we observed more widespread *lacZ* expression and larger individual viral plaques compared to sham-injected controls (*d, e* in the figure). Biologically, we used parasitization by the wasp *Camponotus sonorensis* (Hymenoptera: Ichneumonidae) to immunosuppress *H. zea* larvae. The female reproductive tract of this endoparasitoid contains a symbiotic polydnavirus that is injected during oviposition. In combination with other factors from the wasp, polydnal gene products prevent recognition, encapsulation and destruction of parasitoid eggs and larvae⁷⁻⁹. In our experiments, we allowed *C. sonorensis* to oviposit into *H. zea* larvae immediately before we orally inoculated them with AcMNPV-hsp70/*lacZ*. We subsequently compared *lacZ* expression in these caterpillars with unparasitized control insects and found that the distribution of the reporter gene signal was greater in parasitized *H. zea* than in control insects (*f* in the figure), indicating that the cellular immune response is a significant factor in preventing the spread of infection within *H. zea* larvae.

Our observations on baculovirus infections in the resistant host, *H. zea*, indicate that the immune response defines, in part, the functional host range of AcMNPV and suggests a possible strategy whereby this baculovirus can be genetically manipulated to become a more efficacious pesticide. Specifically, the incorporation of polydnal genes with immunosuppressant activity into the genome of AcMNPV might lower resistance in *H. zea* and other pest species and enable these pests to be controlled with a recombinant of AcMNPV.

**Jan O. Washburn, Bruce A. Kirkpatrick
Loy E. Volkman**

Department of Plant and Microbial Biology,
University of California,
Berkeley, California 94720-3102, USA



Cellular immune response and *lacZ* expression in *H. zea* following inoculation of newly moulted fourth-instar larvae with 44 (*a–e*) or 34 (*f*) polyhedra of AcMNPV-hsp70/*lacZ*. *a*, Colocalization of *lacZ* (blue) and melanization (brown) on the larval midgut at 48 h.p.i. (bar in *a*, 0.37 mm); *b*, melanization (*m*) along tracheal branches expressing *lacZ* at 48 h.p.i. (bar, 0.18 mm); *c*, aggregated haemocytes (*h*) surrounding a melanized capsule (*c*) on a tracheal branch (bar, 100 µm); *d*, maximum distribution of *lacZ* on midgut tracheae from a larval cohort sampled 24 h after intrahaemocoelic inoculation of 1 µl ddH₂O at 18–20 h.p.i. (bar, 0.43 mm); *e*, maximum distribution of *lacZ* on midgut tracheae from a larval cohort sampled 24 h after intrahaemocoelic inoculation of a 1-µl aqueous solution containing 0.1 mg DDCA (Sigma) at 18–20 h.p.i. (bar, 0.43 mm); *f*, maximum distribution at 96 h.p.i. of *lacZ* from a larval cohort parasitized by *C. sonorensis* (below) and a control cohort (above) (bar, 60 mm). In the control larva, blue coloration along the dorsal aorta (upper and lower margin of the whole mount) is an endogenous signal and does not reflect viral infection⁴. In this specimen, viral *lacZ* signals are restricted to much smaller plaques (*p*) in the central region of the midgut.

- Chapman, R. F. *The Insects, Structure and Function* 796–804 (Harvard Univ. Press, 1982).
- Engelhard, E. K. *et al. Proc. Natl Acad. Sci. USA* **91**, 3224–3227 (1994).
- Vail, P. V. *et al. J. Econ. Entomol.* **71**, 293–296 (1978).
- Washburn, J. O. *et al. Virology* **91**, 561–568 (1995).
- Engelhard, E. K. & Volkman, L. E. *Virology* **209**, 384–389 (1995).
- Nobel, C. S. I. *et al. J. Biol. Chem.* **270**, 26202–26208 (1995).
- Summers, M. D. & Dib-Hajj, S. D. *Proc. Natl Acad. Sci. USA* **92**, 29–36 (1995).
- Strand, M. R. & Pech, L. L. *Annu. Rev. Entomol.* **40**, 31–56 (1995).
- Webb, B. A. *Arch. Insect Biochem. Phys.* **26**, 147–163 (1994).

HIV-1 subtype and second-receptor use

SIR — The identification of the chemokine receptors fusin (now renamed CXCR4) and CCR5 as the main co-receptors for HIV-1 entry into CD4⁺ cells of the immune system will assist our understanding of HIV-1 transmission and pathogenesis¹⁻⁶. Here we report data showing that the phenotype of the virus, rather than its genetic subtype, determines receptor usage.

The phenotype of a virus is closely associated with whether it preferentially uses CXCR4 or CCR5: NSI viruses, which are macrophage-tropic and do not

is also a more divergent set of viruses known as the outlier (O) group. It is possible that different genetic subtypes might preferentially use different co-receptors, hence contributing to the uneven worldwide distribution of viral subtypes. The only recent report on this topic indicates that virus strains from subtypes A, C and E can use CCR5, although only a few viruses were tested⁵.

We therefore examined 18 viral isolates from the HIV-1 M group (at least three for each of subtypes A, B, C, D and E), and two from the O group, for second-

tion may not be at the expense of losing the ability to use CCR5. Alternatively, it is possible that most SI isolates contain a mixed population of SI and NSI quasi-species, and thus seem to be able to use both CCR5 and CXCR4.

To confirm our observations on HOS cells, we tested the ability of isolates from different subtypes to replicate in peripheral blood mononucleocytes (PBMC) from individuals homozygous or heterozygous for a recently identified defective allele of CCR5 that contains a 32-nucleotide deletion (Δ 32) in the second external loop and encodes a non-functional protein for HIV-1 entry^{7,8}. We reasoned that if isolates from non-B subtypes could use cofactors other than CCR5 for entry into PBMC, they would replicate in the cells of Δ 32 homozygotes. We therefore tested the same panel of viruses for replication in stimulated PBMC from two homozygous (Δ 32/ Δ 32), one heterozygous (wt/ Δ 32) and one wild-type (wt/wt) individuals (see table). No NSI isolate, irrespective of the genetic subtype, could replicate in PBMC from Δ 32 homozygotes, but all replicated efficiently in wild-type and heterozygote PBMC. In contrast, every SI isolate could replicate in all PBMC regardless of CCR5 genotype, presumably by using CXCR4. Thus, there is an absolute requirement for the presence of a functional CCR5 allele for replication of NSI strains of all the genetic subtypes tested.

We conclude that CCR5 and CXCR4 are the main co-receptors for HIV-1 subtypes A-E and group O. Viral phenotype is dominant over genotype in terms of co-receptor usage. Thus, the uneven worldwide distribution of HIV-1 subtypes is more likely to be the result of stochastic dissemination. However, it is conceivable that sequence polymorphisms in CCR5 and CXCR4 in different human populations could affect the efficiency of HIV-1 entry in a subtype-dependent manner. Finally, the common usage of CCR5 and CXCR4 by multiple genetic subtypes suggests that drugs targeting these receptors might not be unduly compromised by HIV-1 genetic diversity.

Linqi Zhang
Yaoping Huang

Tian He

Yunzhen Cao

David D. Ho*

Aaron Diamond AIDS Research Center,
The Rockefeller University,
New York,
New York 10016, USA
e-mail: dho@adarc.org

*To whom correspondence should be addressed.

1. Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. *Science* **272**, 872-877 (1996).
2. Deng, H. et al. *Nature* **381**, 661-666 (1996).
3. Dragic, T. et al. *Nature* **381**, 667-673 (1996).
4. Doranz, B. J. et al. *Cell* **85**, 1149-1158 (1996).
5. Choe, H. et al. *Cell* **85**, 1135-1148 (1996).
6. Alkhatib, G. et al. *Science* **272**, 1955-1958 (1996).
7. Liu, R. et al. *Cell* **86**, 367-377 (1996).
8. Samson, M. et al. *Nature* **382**, 722-725 (1996).

SECOND-RECEPTOR USAGE BY DIFFERENT GENETIC SUBTYPES OF HIV-1

Virus	Subtype	Phenotype	p24 production in HOS.CD4 cells*				p24 production in PBMC		
			CCR1-4 [†]	CCR5	CXCR4	PB [‡]	wt/wt	wt/ Δ 32	Δ 32/ Δ 32 [§]
93KE101	A	NSI	-	+	-	-	+	+	-
93IN103	A	NSI	-	+	-	-	+	+	-
92RW009	A	SI	-	+	+	-	+	+	+
92HA593	B	SI	-	+	+	-	+	+	+
92HA594	B	SI	-	+	+	-	+	+	+
92HA596	B	SI	-	+	+	-	+	+	+
JRCSF	B	NSI	-	+	-	-	+	+	-
JRFL	B	NSI	-	+	-	-	+	+	-
NL43	B	SI	-	-	+	-	+	+	+
92ZW101	C	NSI	-	+	-	-	+	+	-
92ZW102	C	NSI	-	+	-	-	+	+	-
92ZW106	C	SI	-	-	+	-	+	+	+
94KE102	D	NSI	-	+	-	-	+	+	-
94KE103	D	NSI	-	+	-	-	+	+	-
92UG046	D	SI	-	+	+	-	+	+	+
93TH304	E	SI	-	+	+	-	+	+	+
93TH305	E	NSI	-	+	-	-	+	+	-
93TH307	E	NSI	-	+	-	-	+	+	-
CA9	O	NSI	-	+	-	-	+	+	-
MVP5180	O	SI	-	+	+	-	+	+	+

* Plus symbol, p24 concentration > 60 pg ml⁻¹; minus symbol, $\leq$ 10 pg ml⁻¹.

[†] Identical results were obtained with CCR1, 2, 3 and 4.

[‡] Negative control cells containing vector pBABE-puro only².

[§] Identical results were obtained with PBMC from two cases.

induce syncytia, use CCR5, whereas SI strains, which induce syncytia or are adapted to T-cell lines, prefer CXCR4. There is one example of a broadly tropic HIV-1 strain able to use both CCR5 and CXCR4 for entry⁴. In addition, CCR1, CCR2b and CCR3 can function to a limited extent as HIV-1 co-receptors for some viral strains³⁻⁵. With a single exception⁵, these studies used HIV-1 strains or *env* genes derived solely from genetic subtype B, the subtype of HIV-1 that predominates in North America and Europe but not globally. Nine HIV-1 genetic subtypes (A to I) have now been designated within the major (M) group, and there

receptor usage. These 20 HIV-1 isolates, originating from 8 different countries, had previously been genetically subtyped and their NSI or SI phenotypes determined in MT-2 cells. To assess second-receptor use, we added 500 TCID₅₀ (half-maximal tissue-culture infectious dose) of each isolate to HOS cells engineered to express human CD4 and CCR1, CCR2b, CCR3, CCR4, CCR5 or CXCR4 (ref. 2). We assessed virus replication using supernatant p24 antigen concentration on day 6 after infection.

We found that each NSI isolate could utilize CCR5 for entry, irrespective of its genetic subtype (see table). However, SI isolates could use both CCR5 and CXCR4, except for one subtype B TCLA virus (NL43) and one subtype C SI primary isolate (92ZW106), which could only utilize CXCR4. The finding that dual-receptor usage is more common among SI strains suggests that adaptation of HIV-1 to use CXCR4 during phenotypic evolu-

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. Priority will be given to letters of fewer than 500 words and ten references.

Noise in human muscle spindles

SIR — Stochastic resonance is a phenomenon in which the response of a non-linear system to a weak input signal is optimized by the presence of a particular non-zero level of noise^{1,2}. It has been demonstrated in several biological systems, ranging from ion channels³ to sensory neurons⁴ to human psychophysics⁵. Here we demonstrate stochastic resonance *in vivo* in the human proprioceptive system. We show that the sensitivity of muscle-spindle receptors to a weak movement signal can be maximally enhanced by introducing noise through the tendon of the parent muscle.

We recorded firing activity of individual muscle-spindle afferents from the wrist and hand extensor muscles from the radial nerve in healthy human subjects. When a suitable afferent is isolated, the subject's right wrist is passively rotated by a manipulandum, with a low-amplitude sinusoidal waveform (Fig. 1). We applied a random noise input by a tendon stimulator to stretch the appropriate muscle, and varied the intensity of the noise input between trials.

To characterize stochastic resonance in the spindle afferents, we calculated the output signal-to-noise ratio from the power spectrum, defining it as the ratio of the strength (area) of the signal peak to the mean amplitude of the back-

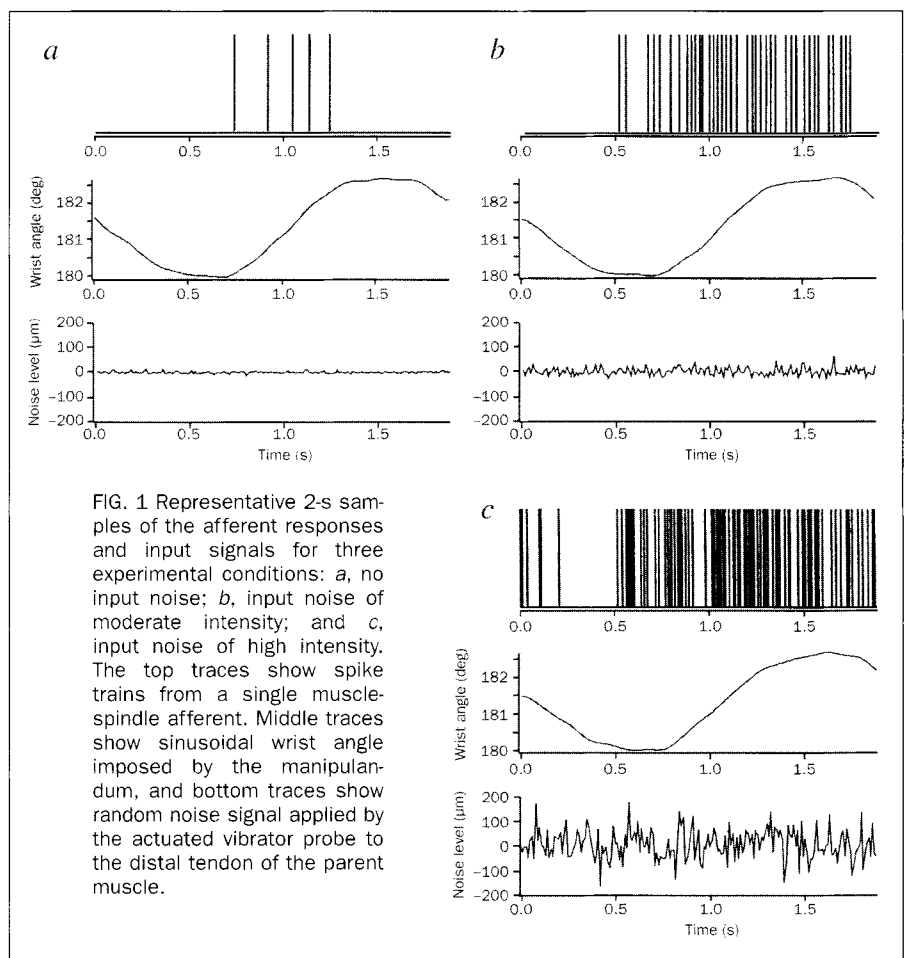


FIG. 1 Representative 2-s samples of the afferent responses and input signals for three experimental conditions: *a*, no input noise; *b*, input noise of moderate intensity; and *c*, input noise of high intensity. The top traces show spike trains from a single muscle-spindle afferent. Middle traces show sinusoidal wrist angle imposed by the manipulandum, and bottom traces show random noise signal applied by the actuated vibrator probe to the distal tendon of the parent muscle.

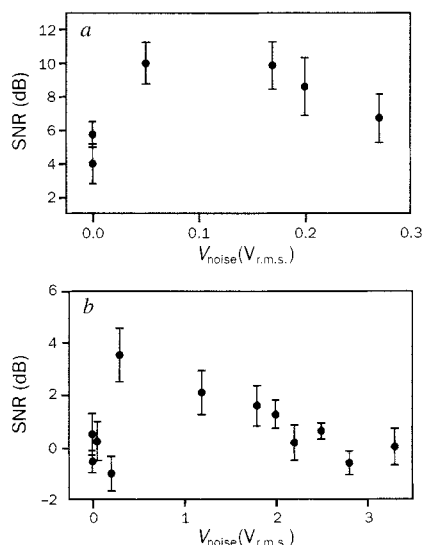


FIG. 2 Values of the output signal-to-noise ratio (SNR) versus the input noise r.m.s. voltage for single spindle afferents in two different subjects: *a*, a healthy subject; *b*, a subject who was administered an anaesthetic block of the brachial plexus. The output SNR for each noise intensity level was computed from the spike-train power spectra in 1-Hz increments. Error bars on the SNR values represent the standard deviation of the noise value of the centre frequency of each 1-Hz bin.

ground noise at the input signal frequency. The output signal-to-noise assesses the coherence of the system response (its spiking activity) with the input signal at the frequency of wrist rotation.

Six of the eight afferents we examined showed clear stochastic resonance behaviour: as input noise intensity increased, the output signal-to-noise ratio rapidly increased to a peak and then slowly decreased (Figs 1 and 2). Thus, in each of these cases, the presence of a particular, non-zero level of noise significantly enhanced the sensitivity of the muscle spindle receptor to the weak input signal (see also ref. 6).

Stochastic resonance from mechanical stimulation of the arm could occur directly on the muscle-spindle receptor or indirectly through cutaneous fusimotor reflexes^{7,8}, because the noise input presumably activates cutaneous receptors beneath the probe as well as muscle spindles in the muscle proper. To control for a possible cutaneous fusimotor reflex, we inactivated the fusimotor system in the right arm of one subject with an anaesthetic block of the brachial plexus, functionally denervating the arm at the shoulder. The afferent recordings from that subject exhibited clear stochastic resonance characteristics (Fig. 2*b*), suggesting that the positive effect of

noise reported in this study is due to a direct effect on the muscle-spindle receptor rather than as a result of cutaneous fusimotor reflexes.

Our results raise the possibility that the fusimotor system uses a stochastic resonance-type mechanism to increase the sensitivity of muscle spindles to stretch. It is known that fusimotor activity and muscle spindle output is enhanced during postural tasks and novel movements⁹. Fusimotor activation, which increases the random properties of muscle spindle firing¹⁰, could provide the stochastic input required by a stochastic resonance mechanism. Thus, fusimotor neurons may serve as a beneficial,

1. Benzi, R., Sutera, A. & Vulpiani, A. *J. Phys. A* **14**, L453–L457 (1981).
2. Moss, F., Pierson, D. & O'Gorman, D. *Int. J. Bifurc. Chaos* **4**, 1383–1397 (1994).
3. Bezrukov, S. M. & Vodyanov, I. *Nature* **378**, 362–364 (1995).
4. Douglass, J. K., Wilkens, L., Pantazidou, E. & Moss, F. *Nature* **365**, 337–340 (1993).
5. Collins, J. J., Imhoff, T. T. & Grigg, P. *Nature* **383**, 770 (1996).
6. Matthews, P. B. C. & Watson, D. G. *J. Physiol. (Lond.)* **317**, 365–381 (1981).
7. Johansson, H. & Sojka, P. *J. Physiol. (Lond.)* **366**, 343–363 (1985).
8. Gandevia, S. C., Wilson, L., Cordo, P. J. & Burke, D. *J. Physiol.* **479**, 499–508 (1994).
9. Prochazka, A., Hulliger, M., Zangger, P. & Appenteng, K. *Brain Res.* **339**, 136–140 (1985).
10. Matthews, P. B. C. & Stein, R. B. *J. Physiol. (Lond.)* **202**, 59–82 (1969).

intrinsic noise source for the proprioceptive system.

Paul Cordo

Robert S. Dow Neurological Sciences
Institute,

Portland, Oregon 97209, USA

email: cordo@lhs.org

**J. Timothy Inglis*, Sabine Verschuereen,
James J. Collins, Daniel M. Merfeld,
Stuart Rosenblum, Scott Buckley,
Frank Moss**

*Addresses: University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada (J.T.I.); School of Human Kinetics, Department of Kinesiology, Catholic University of Leuven, Heverlee, B3001 Belgium (S.V.); NeuroMuscular Research Center & Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215, USA (J.J.C.); Robert S. Dow Neurological Sciences Institute, Portland, Oregon 97209, USA (D.M.M. & S.B.); Department of Anesthesiology, Legacy Emanuel Hospital and Medical Center, Portland, Oregon 97227, USA (S.R.); Department of Physics and Astronomy, University of Missouri, St Louis, Missouri 63121, USA (F.M.).

Noise-enhanced tactile sensation

SIR — Stochastic resonance, a phenomenon in which noise enhances the response of a nonlinear system to a weak signal^{1,2}, has been demonstrated to be important functionally in various physiological systems³⁻⁶. Here we demonstrate improvements in human sensory perception mediated by stochastic resonance-type effects. We show that the ability of an individual to detect a subthreshold tactile

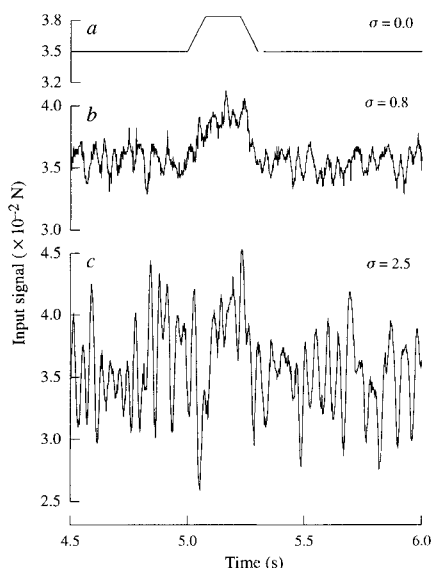


FIG. 1 Representative samples of the input signal for three experimental conditions: a, stimulus with no input noise; b, stimulus with input noise of moderate intensity; c, stimulus with input noise of high intensity. (Respective values for the input noise standard deviation σ are in units of 10^{-3} N.) The amplitude of the stimulus was adjusted so that the stimulus was just subthreshold. The input noise consisted of zero-mean gaussian 'quasi-white' noise. The input signal was convolved with a low-pass filter (cutoff frequency 30 Hz) to limit the excitation of rapidly adapting afferents.

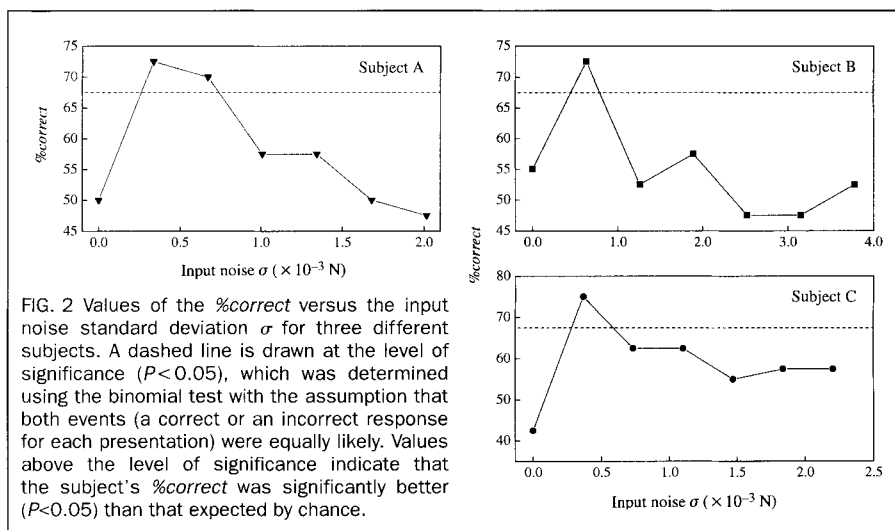


FIG. 2 Values of the %correct versus the input noise standard deviation σ for three different subjects. A dashed line is drawn at the level of significance ($P < 0.05$), which was determined using the binomial test with the assumption that both events (a correct or an incorrect response for each presentation) were equally likely. Values above the level of significance indicate that the subject's %correct was significantly better ($P < 0.05$) than that expected by chance.

stimulus can be significantly enhanced by a particular level of noise.

We conducted a series of psychophysical experiments on 10 healthy young subjects. We applied local indentations to the tip of each subject's right middle digit. The protocol consisted of the presentation of either a subthreshold stimulus plus noise (Fig. 1) or no stimulus plus noise, and we instructed the subjects to indicate when they detected a stimulus. Each trial consisted of 20 presentations, equally distributed between 'stimulus' and 'no stimulus', with a randomized presentation sequence and an interpresentation interval of 5 s. We held the intensity of the input noise constant for each trial and varied it between trials (Fig. 1); we included 7–9 different noise intensity levels in the protocol and conducted two trials for each noise level in randomized order.

To characterize stochastic resonance-type behaviour, we used a measure, %correct, that quantifies the percentage of trials for which a subject correctly identifies the presentation of 'stimulus' or 'no stimulus'. The %correct should, on average, be 50 for a protocol involving a subthreshold stimulus and an equal number of 'stimulus' and 'no stimulus' presentations. On the other hand, this measure should be near 100 for a protocol with test stimuli that are well above the detection threshold. Nine of the ten subjects we examined exhibited clear stochastic resonance-type behaviour: as input noise intensity increases, the %correct increases significantly to a peak ($P < 0.05$) and then decreases (Fig. 2). We retested several of these subjects on subsequent days and obtained similar results. In one of the ten cases, the introduction of input noise did not significantly affect the subject's ability to detect the subthreshold stimulus.

These findings indicate that input noise can serve as a 'negative masker' for subthreshold tactile stimuli, that is, noise can increase the detectability of weak sig-

nals. Negative masking has been observed in vibrotactile for cases where the test stimulus and the masker (or pedestal) are sinusoidal signals of the same frequency and phase⁷⁻⁹. This effect has been shown to be robust to small levels of background noise, provided the sinusoidal pedestal is present⁹. We have shown here that, under certain conditions, the noise itself can be used as a suitable pedestal for enhancing the detection of a subthreshold stimulus.

These results suggest that a noise-based technique could be used to improve tactile sensation in humans. Such a technique could be incorporated into the design of haptic interfaces for telerobotics and virtual environments. From a clinical standpoint, a stochastic resonance-based technique could be applied to individuals with elevated cutaneous sensory thresholds, such as older adults¹⁰ and patients with peripheral neuropathies¹¹ or strokes¹².

James J. Collins, Thomas T. Imhoff

NeuroMuscular Research Center and

Department of Biomedical Engineering,

Boston University,

Boston, Massachusetts 02215, USA

Peter Grigg

Department of Physiology,

University of Massachusetts Medical

Center,

Worcester, Massachusetts 01655, USA

- Wiesenfeld, K. & Moss, F. *Nature* **373**, 33–36 (1995).
- Collins, J. J., Chow, C. C. & Imhoff, T. T. *Nature* **376**, 236–238 (1995).
- Douglass, J. K., Wilkens, L., Pantazidou, E. & Moss, F. *Nature* **365**, 337–340 (1993).
- Levin, J. E. & Miller, J. P. *Nature* **380**, 165–168 (1996).
- Collins, J. J., Imhoff, T. T. & Grigg, P. J. *Neurophysiol.* **76**, 642–645 (1996).
- Cordo, P. J. *et al.* *Nature* **383**, 769–770 (1996).
- Hamer, R. D., Verrillo, R. T. & Zwislöck, J. J. *J. Acoust. Soc. Am.* **73**, 1293–1303 (1983).
- Verrillo, R. T., Gescheider, G. A., Calman, B. G. & Van Doren, C. L. *Percept. Psychophys.* **33**, 379–387 (1983).
- Gescheider, G. A., Verrillo, R. T. & Pelli, D. G. *J. Acoust. Soc. Am.* **91**, 348–353 (1992).
- Verrillo, R. T. *Sensory Processes* **3**, 49–59 (1979).
- Gelber, D. A. *et al.* *J. Diabetes Complications* **9**, 170–176 (1995).
- Robertson, S. L. & Jones, L. A. *Arch. Phys. Med. Rehabil.* **75**, 1108–1117 (1994).

The gospel of contingency

Christian de Duve

Full House: The Spread of Excellence from Plato to Darwin. By Stephen Jay Gould. Harmony/Cape. Pp. 244. \$25, £16.99. UK title is "Life's Grandeur".

WHAT is the difference between a drunkard who, staggering out of a pub, ends in the gutter, and life, which, starting from primitive bacteria, eventually came to compose the "Well-tempered Clavier", paint the Sistine Chapel ceiling and discover natural selection? None, according to the distinguished palaeontologist and popular science writer Stephen Jay Gould. In both cases, random movement devoid of directional bias becomes cumulatively slanted in one direction — either the gutter or humankind — for the simple reason that any excursion in the other direction bumps into a wall — either the pub's wall or the limit of minimal complexity compatible with life. Only one direction is open and, because of the laws of probability, is invaded ever farther, given enough time.

Other 'case studies' are brought up in support of the book's main theme — consider the 'full house', not just a small part of it. These include Gould's own successful bout with mesothelioma, a moving story, told with commendable restraint, of the author's finding strength — fortunately helped by an effective drug — in his discovery that the survival distribution curve of patients afflicted with the disease has an extended right tail; the evolution of the horse, a false paradigm of directionality merely expressing the failure of all other branches to pass the natural-selection test; the increase in size of some planktonic foraminifera after each major extinction, a typical case of a random walk in the only direction open to it; the evolutionary differentiation of vertebrae, with higher and lower complexity about equally favoured; and, in between, a long (a quarter of the book) and tedious analysis of why baseball hitters no longer achieve a 0.400 batting average, a problem whose solution "stunned and delighted (beyond all measure)" the author "by the elegance and clarity of the result".

The book ends with an ode to the mighty bacteria, which have been around for more than 3.5 billion years, occupy every possible niche under the sun or away from it, and still make up the mode in the distribution of living forms, even beating, according to some shaky calculations, the

combined trees of all our forests in total biomass (a strange measure of importance, incidentally). We humans, on the other hand, are no more than the puny extreme of a distribution curve that has gradually become skewed in the direction of complexity by the 'drunkard's random walk' of evolution. We are, despite the intuition of



our misguided hubris, no more than a tiny tail ridiculously trying to wag a large dog. If you believe otherwise, try living in boiling water, in the Arctic Ocean, in drying brine or inside deeply buried rocks.

To those familiar with Gould's ideas, this message will come as no news. What will surprise is the tone of the book, which is very different from the author's elegantly written columns in *Natural History*, the monthly magazine of the New York Museum of Natural History, and from such masterly displays of lucid, scholarly erudition as his *Ontogeny and Phylogeny* (Harvard University Press, 1977). Abandoning sober intellectual discourse and relying instead on parables and assertions to spread the gospel of contingency, Gould here presents his ideological views in characteristic evangelical style, admonishing all of us sinners to mend our ways and see the

light, as he himself has done before us, vituperating against all those who persist in their errors and fail to recognize the truth.

Indeed, much of the book is devoted to denouncing common errors and fallacies, nouns that, together with such variations as delusion, mistake, misconception, misconstruction, misreading, misunderstanding, prejudice, bias, distortion, wrongness and myopic vision, and the corresponding adjectives, occur some 150 times — I made a rough count — throughout the work, with, as counterweight, the author's own "insight", his "eureka" that "humans can occupy no preferred status as a pinnacle or culmination", his discovery that "life has always been dominated by the bacterial mode" and always will be.

For that is the pernicious heresy that needs to be uprooted, the arrogantly anthropocentric belief that the emergence of humankind and, more generally, of consciousness and intelligence, reflects some kind of directionality in biological evolution. The many subtle ways in which this concept of directionality can be dissected and the searching analyses to which it may be, and has been, subjected in the light of modern knowledge are all disregarded in favour of the sophist's strategy of building a straw man, in which complexity, an objective reality that the author recognizes, albeit reluctantly, is insidiously converted into the subjective notion of progress. This notion is then attacked in a manner more suited to the end of the nineteenth century — the author's favourite — than to that of the twentieth. Indeed, except for a few references to some recently discovered bacteria, Gould ignores all the late developments in genetics, biochemistry, cell biology and molecular biology, keeping strictly to his own turf of descriptive palaeontology and choosing astonishingly trivial examples to support his message. Nowhere is a case of true evolutionary increase in complexity addressed in any serious fashion. Nowhere are the molecular mechanisms and constraints of Darwinian evolution mentioned, let alone taken into account.

The book is clearly aimed at a new target for Gould: a readership that needs to be told that living organisms do not disobey the second law of thermodynamics and why; that has yet to be introduced to elementary notions of probability and to the shapes of statistical distribution curves; that must be taught the difference between a mean, a median and a mode, and shown the lack of significance of a mean without its variance; that is more interested in baseball than in evolution, but can be lured by

love of the former to take an interest in the latter and so become converted to Gould's creed of contingency. If you don't recognize yourself in this description, don't read the book, especially if you are a Gould aficionado. It will only make you sad. □

Christian de Duve, emeritus professor at the University of Louvain, Belgium, and at the Rockefeller University, New York, is at the International Institute of Cellular and Molecular Pathology, 75 Avenue Hippocrate, B-1200 Brussels, Belgium. His most recent book, Vital Dust: Life as a Cosmic Imperative, is now out in paperback. BasicBooks, \$14.

Why God plays dice

David Mermin

Beyond Science. By John Polkinghorne. Cambridge University Press: 1996. Pp. 131. £13.95, \$19.95.

ALL those with even a modicum of intelligence who believe that 'creation science' should be given equal time with evolution in school curricula ought to read this charming little book. John Polkinghorne makes a case that, far from being at odds with the existence of a creator, the fact that evolution is possible at all is powerful evidence of underlying design at another level.

A prominent role in the argument is played by the constellation of ideas associated with the various anthropic principles. Evolution points to God as the Great Tuner of fundamental constants. Polkinghorne finds this far more compelling than the notion that our Universe appears delicately adjusted to make the evolution of life possible, because if it weren't we wouldn't be here to notice it. And he is sceptical of the view that, if the stability of atomic nuclei were grossly altered by tiny changes in certain coupling constants, different avenues would open up to the development of complex carriers of intelligence, which we simply lack the wisdom to imagine.

In addition to the miraculous values of the fundamental constants, there is also evidence of design in the wonderful "opportunity for the interplay of chance and necessity" afforded by quantum theory, which seems ingeniously contrived to be both flexible enough to allow evolutionary variation, yet not so floppy as to threaten the persistence of successful new life forms.

Polkinghorne's grand extension of the argument from design shifts the grounds for scepticism from Darwin back to Hume, about whom he has only a little to say. The problem of evil, he remarks, "can at least be addressed by the insight that

this is the necessary cost of a universe allowed to make itself, whose shuffling explorations of possibility will have to have ragged edges". The question of who tuned the Tuner does not come up.

Yet another facet of his case is provided by the mystery of consciousness: why should natural selection have given rise to self-awareness? Would not mindless automata have been at least as successful? The views of many philosophers, here and elsewhere, receive some well-deserved jabs. (On exploring the nature of consciousness by way of thought experiments with extraordinary duplicating machines: "Well, philosophy is wonderful, but peculiar premises may lead to peculiar conclusions.") So do computer models of the mind, for computers are useless without a program. In Polkinghorne's view of consciousness, God appears to be the Prime Programmer Unprogrammed.

Only apparently unrelated to these religious concerns is a critical examination, early in the book, of the account of science as a social construction. This contribution to the 'science wars', unusually temperate for a scientist, lucidly states the claims and articulates the naiveties of the so-called 'strong programme' in the sociology of scientific knowledge. Polkinghorne can be poetic in his defence of objective reality: "Far from the physical world proving to be like clay in our theoretical hands, it displays a diamond-like hardness, resistant to our expectations and imposing upon our minds its idiosyncratic and unanticipated structure." Scientists who admire this elegant dismantling of the view that their discipline lacks objective content may be startled to find later in the book the same thoughtful approach applied to the view that moral and aesthetic principles are social constructions. I suspect the defence of the objectivity of scientific knowledge may have been cunningly contrived to set us up for his attack on cultural relativism.

I have two major criticisms. Polkinghorne takes our ability to treat nature at the quantum level as evidence that our capabilities go well beyond anything evolution could have given rise to, for nothing in the struggle for survival could have required us to be creatures capable of such an understanding. I would have put it just the opposite way. It is because nothing required us to apprehend atomic structure during our evolutionary development that we are incapable of understanding what it is that quantum physics describes. Quantum mechanics is weird to us because we can make inferences about the atomic world only indirectly through the correlations we can arrange for it (called measurements) with those parts of the world (called classical) that evolution has outfitted us directly to apprehend. Polkinghorne dismisses too lightly the

mysterious character of quantum mechanics. Although I think I know what he has in mind when he says "[y]ou could never build a wave out of finite collections of particles, but a wave-like state is one with an indefinite number of particles making it up", I would not agree that this explains "how the trick is done" in electron diffraction.

And an eloquent discussion of how it may all end — decay into low-grade radiation, fiery Big Crunch or the persistence of ever-adapting forms of life to the very end — comes down from these lofty heights with a resounding thud for the non-Christian reader with a final slightly uncomfortable ("candour requires me to add") declaration of faith in the resurrection of Jesus Christ. Provincial Christian mythology is a blemish on so grand a theological vision.

I nevertheless enjoyed the book immensely. Although I believe this late-twentieth-century version of the argument by design carries many of the flaws of its venerable predecessors, Polkinghorne's literate sense of wonder at the magical richness of things shines out on almost every page, whether or not one agrees that it implies a creator. □

David Mermin is in the Department of Physics, Cornell University, Clark Hall, Ithaca, New York 14853-2501, USA.

From the top down

Philip W. Anderson

How Nature Works: The Science of Self-Organized Criticality. By Per Bak. Copernicus/Springer: 1996. Pp. 205. \$27, £20.50.

THE rarefied air of the Santa Fe Institute in New Mexico seems to encourage the universalist persuasion: to each scientist, his great idea really seems to be the "theory of everything" that has escaped previous notice, and supersedes all of those promulgated by previous Santa Fe sages (and others). Per Bak now tries to persuade us that he has uniquely found out "how nature works".

His particular version of the theory of everything is 'self-organized criticality', a theory and a generic scenario of which I have long been a public advocate. It leads, in my mind correctly, to sobering and counterintuitive conclusions about a wide variety of natural and social processes such as climate, tectonics and the macroeconomy. I therefore consider this book a 'must read', despite its exaggerated claims and obvious weaknesses, but because of its importance I could wish that it had been written with more care.

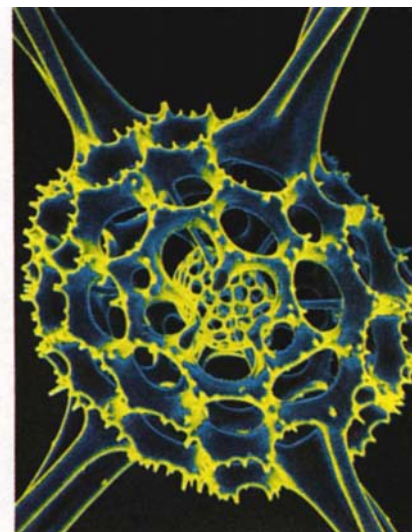
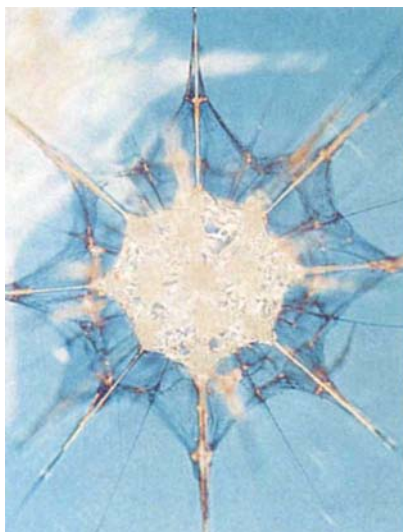
First, let me give a rough idea of Bak's

thinking. He takes off from the observation, ably publicized by Benoit Mandelbrot, of the widespread occurrence in nature of 'scale-free' power-law distributions, to which Mandelbrot gave the fortunate name of 'fractals'. The power law appears also in time series — that is, in fluctuations (where it is often called '1/f noise') — and its main practical consequence is that very large fluctuations cannot, in general, be ignored in favour of the cumulative effect of small ones. Mandelbrot was content to pile example on example, but Bak has taken the considerable and important step of trying to produce a mechanistic explanation of these observations.

Bak makes an analogy between the scale-free, power-law regime near a thermal phase transition or 'critical point' and the 'critical' behaviour of a large dissipative, nonlinear dynamical system driven by some slow, large-scale driving force such as solar energy or the slow motion of tectonic plates. In the former case, he argues, the critical point must be achieved by careful tuning of the external parameters such as temperature and pressure, whereas in the latter the system tunes itself to be very close to what he calls 'criticality' by dint of just barely achieving the state in which there is large-scale motion. In both cases there seems to be no natural scale for the sizes of fluctuations, so we see 'critical fluctuations'; but in the latter these arise automatically without tuning, so are 'self-organized'. In the critical state, large fluctuations appear, which he calls 'avalanches'.

With a sequence of simple computer models — the most convincing of which is a gradually growing sand pile — he demonstrates convincingly the possibility of such behaviour and describes in fascinating detail the process by which he and his isolated little group arrived at them. The first four-and-a-half chapters, where these ideas and their first, most successful applications to geology and to earthquakes appear, are a solid contribution to our thinking about complex systems, although, as Bak admits, arguments for universal behaviour seem not quite to fit all real, practical cases.

The other half of the book is far more speculative, and I, at least, cannot accept all of it at face value as readily as the original 'sand-pile' picture. Here we go on to more genuinely complex systems: first computers, then coevolving ecologies ("coevolving dancing landscapes") and the history of evolution, to the brain, to economics, and what have you. Bak's indubitable talent for finding simple models with marvellously complex behaviour comes to the fore, and indeed he can model 'extinction avalanches' in a model of an ecology; but the question here becomes whether the model really captures the essence of the phenomenon or is constructed simply to do what it does. An evolving ecology, or a brain, has more to it



WHY are eggs egg-shaped and fish fish-shaped? Why do planets look like balls rather than squares or pyramids? Mathematicians Stefan Hildebrandt and Anthony Tromba look at the centuries-old search for fundamental laws governing nature's design schemes. Their book, *The Parsimonious Universe: Shape and Form in the Natural World*, draws on examples from astronomy to microscopy, including these radiolarians — living (left) and skeletal. Springer, \$32, £19.50.

than the right statistical distribution of extinctions or of firings. Contrary to Stephen Jay Gould's prejudice, evolution does have a direction (we need not call it 'progress'), and it is not a static replacement of one species by another. So Bak's 'life' is not life as we know it.

Homeostasis, indeed, is a principal ingredient of the workings of all truly complex systems, and Bak's ideas are important in telling us how homeostasis works. But they are not all there is to a brain, an ecology or an economy. Nonetheless, there is much meat in these chapters. The basic message, that large fluctuations (avalanches, storms, depressions, earthquakes) are vital to the dynamics of large systems, is an important and widely ignored fact.

Bak writes with such ease and lucidity, and his ideas are so intriguing, that one reads along only hoping that all is to be believed and accepted at face value. Yet again and again, in those parts of the book that can be personally verified, one regularly finds dropped clangers. Perhaps the intellectual history of chaos is not too important (ascribed almost entirely to Mitch Feigenbaum, a judgement with which only Feigenbaum would concur) and even less so is the history of Brookhaven Laboratory's condensed-matter-physics group. Why does Bak spend a page and a half extolling Brookhaven for its many Nobel laureates in high-energy physics, and then practically without a break set out bad-mouthing the entire enterprise of particle physics? Such thoughtless writing and venting of personal prejudices alert one to possibly more serious omissions.

Another minor error in a field I know well has to do with pulsar glitches. In this connection Bak quotes two authors from Ilya Prigogine's institute who seem to have revived (20 years later and without attribution) the Pines-Shaham starquake theory, which its authors long since abandoned in favour of our mutual work on vortex jumps in the superfluid core. To be ignorant of a decade's work by two of his Santa Fe colleagues is indicative of Bak's scholarship.

Perhaps more serious is his failure to discuss the entire field of hydrodynamics which has served as the classic model for dissipative dynamics. His arguments for self-organized criticality sound much like a generalization of the 50-year-old arguments for scaling (power) laws in fully developed turbulence. It is disingenuous for the book to contain no attempt to relate the two subjects, even while remarking on the fractal outcome of hydrodynamics in the form of weather. In particular, turbulence shows one way in which the arguments can fail, namely qualitative inhomogeneity of the system.

Another missing discussion is of alternative mechanisms for the various Fick's laws and for power laws in economics.

Nevertheless, this book is essential reading for those interested in complex systems in general. Its idiosyncratic personal style may intrigue as many as it turns off, but, like the similarly flawed classic book of Mandelbrot, it will reward a sufficiently sceptical reader. □

Philip W. Anderson is at the Joseph Henry Laboratories of Physics, Princeton University, Princeton, New Jersey 08544, USA.

Cranks 0, Schoolboys 1

F. E. A. Johnson

Fermat's Last Theorem. By Amir D. Aczel. *Four Walls Eight Windows*: 1996. Pp. 147. \$18, £12.99.

IN his public life, Pierre de Fermat was a successful lawyer, in his private life a mathematician. He published little, and was wont to jot down his discoveries as marginal notes. Some time in the 1630s, in his copy of Diophantus's *Arithmetic*, opposite a discussion of the solutions of the



Fermat: puzzle took centuries to crack.

equation $x^2 + y^2 = z^2$ (which rises, of course, in connection with Pythagoras's theorem), Fermat observed that, by contrast, the only integer solutions to the corresponding equation $x^n + y^n = z^n$, with $n \geq 3$, are the trivial ones, in which some term is zero. This is Fermat's last theorem. Fermat provided no proof, and thereafter generations of mathematicians tried, but failed, to find one. The problem became notorious for attracting cranks (and schoolboys), and the general view among mathematicians was that, in common with the flawed attempts in the nineteenth century by Lamé and Kummer, Fermat's presumed 'proof' was fallacious. Just over two years ago, however, the English mathematician Andrew Wiles seems finally to have produced a correct proof.

This is a book of two distinct halves. The author knows all the old stories recounted by Eric Temple Bell in *Men of Mathematics* (1937), and in the first half, dealing with the history, cannot resist retelling quite a few of them, regardless of their strict relevance to the matter in hand. Galois's duel, Cauchy's shameful treatment of Abel, even Euler and the

Konigsberg bridges are all trotted out. This is wonderful material, but the author fails to do it justice. In fact, he is on a hiding to nothing. In a book of this length, he cannot hope to compete with Bell's popular classic, which despite its sententious tone is still without a serious rival.

The author's approach is chatty rather than precise, and I found enough minor inaccuracies in what was familiar (Mordell was American, not English; the Bolyai-Lobachevsky geometry on page 70 should be hyperbolic, not spherical) to give me qualms that the unfamiliar, although broadly correct, might not literally be so. For example, detecting a certain tendency towards 'imaginative reconstruction' of inessentials, I found myself worrying whether, in the conversation between Ribet and Mazur, Ribet was actually sipping cappuccino. Perhaps it was espresso. No matter, but it is that sort of book.

However, it is precisely at this point, in the final 44 pages, that the author comes into his own. There are broadly three separate phases. In the first, dealing with the priority dispute over what came to be known as the Taniyama-Shimura-Weil conjecture, he plunges headlong into controversy, and expresses himself in the most forthright manner on the rights and wrongs of the case. Some famous names come out of it badly mauled, but whether the reader can entirely appreciate this section without having witnessed the full spate of André Weil's acerbic wit is another matter.

The second deals with the stroke of genius of Gerhard Frey, and its vindication by Ribet. It is simply this: if $a^p + b^p = c^p$ solves the Fermat equation then the elliptic curve

$$y^2 + xy = x^3 + \frac{1-b^p-c^p}{4}x^2 + \frac{b^pc^p}{16}x$$

violates the Taniyama-Shimura conjecture, so that to prove Fermat it suffices to prove Taniyama-Shimura. Alas, the author misses the opportunity of writing down the Frey equation.

Finally, there is the truly dramatic section that describes Wiles's proof of the Taniyama-Shimura conjecture; his seven-year self-imposed silence while developing his breakthrough techniques; his first and, as it turned out, premature announcement; the discovery of the gap in the proof; and the eleventh-hour rectification of the proof, just as he is about to abandon for good his apparently futile research.

Wiles's finished manuscript was published in *Annals of Mathematics* **142**, 443-551 (1995). For those interested in the details, Wiles gives, in a long introduction, his own chronicle of the solution. □

F. E. A. Johnson is in the Department of Mathematics, University College London, Gower Street, London WC1E 6BT, UK.

Thought blocks

Alan Mackay

The Quotable Einstein. Edited by Alice Calaprice. *Princeton University Press*: 1996. Pp. 269. \$16.95, £12.95.

QUOTATIONS are not epigrams, clever one-off ripostes, neither are they simply significant extracts from some important author. Quotations are clichés, blocks of thought that can be used, indeed have been used, to encapsulate and drive home some idea better than the writer himself could manage. "Even ministers they ha'e been kenn'd, A rousing whid at times to vend, an' nail't wi' Scripture", to use one self-referentially.

Alice Calaprice, from two decades of editing the Einstein papers for publication, has collected significant extracts of Einstein's writing and has also inherited a collection of snippets made by Einstein's formidable secretary Helen Dukas, who was a jealous guardian of the reputation of the great man and one of the literary executors who refused me permission to quote an item from Einstein as being "unworthy".

The struggle for the literary inheritance, of which Freeman Dyson includes an interesting account in his foreword, is now amicably settled, and some 550 quotations by and about Einstein are presented for use.

Some have actually been tested as quotations by being quoted. Many others await this test, but all are provided with the most authoritative references to the Einstein archive. Some items come from the recently available private letters of Einstein to his wife, illuminating his personal marital difficulties but not more generally applicable, and some items illustrate the fact that Einstein shared many of the other failings of more ordinary people. Most of the original statements were in German, and only English translations are given here.

This is a useful collection, but the overall impact of 550 items is less than that of a dozen or two that have proved able to get about under their own power. We learn that the most famous quotation of 1921, "*Raffiniert ist der Herr Gott, aber boshaft ist Er nicht*" ("The Lord God is subtle, but malicious he is not"), now engraved in stone at Princeton, was followed by "I have second thoughts. Maybe God is malicious" (to Valentin Bargmann, quoted in *Einstein in America* by J. Sayen; Crown, 1985), but this does not have the eternal impact of Galileo's "*Eppur si muove*". □

Alan Mackay is in the Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK.

Artful communication

John Morton

The Prehistory of the Mind: A Search for the Origins of Art, Religion and Science. By Steven Mithen. *Thames and Hudson: 1996. Pp. 288. £16.95, \$27.50.*

STEVEN Mithen has produced a dramatic reconstruction of how the modern human mind has evolved over the past few million years. He writes with clarity and verve, and I have already dined out on my new-found ability to sort out which species of *Homo* is which and on Mithen's wide-ranging speculations on what properties of mind are needed for basic human activities. Technical discussions are placed in the extensive end-notes and are not allowed to impede our race through prehistory. The core belief is that only archaeology can help us to understand the origins of the modern mind. The archaeology is used to specify what the minds of our ancestors must have been like.

The book starts with an overview of our fossil history. Mithen then picks out some recent writing in cognitive and developmental psychology. His reading here has been spotty and, although some of his informants are put to good use, others are less well integrated. However, he emerges with a framework of specialized intelligences: linguistic, technical, social and natural history. These are used in the analysis that follows of the chimpanzee's mind and those of our forebears.

In this analysis, chimpanzees have general intelligence as well as specialized domains of social intelligence and natural history. Their expertise at finding tools for the job in hand is seen as a manifestation of general intelligence. Their inability to instruct their young in tool use is seen as a manifestation of the lack of communication between social and general intelligence. The same is true for *Homo habilis*, who has been able to manufacture limited tools — owing to the development of technical intelligence — and exploit animal carcasses, but there is no imposition of social information on the tools.

Homo erectus had the ability to manufacture symmetrical hand-axes, demonstrating an advanced technical intelligence. But they ignored other materials such as bone and antler, they had only general-purpose tools, there were no multi-component tools and there was little or no cultural variation. The Neanderthals were a little further advanced technically, but managed to survive in a harsh environment owing to their advanced natural-history intelligence, which enabled them to track and trap game.

Mithen deduces that they must have had high social intelligence because they cooperated but, on the basis of the archaeological evidence, that they had little social structure. The solution to the

paradox is that the technical, natural-history and social intelligences have cognitive barriers between them. Modern humans, with interconnected intelligences, would sit round the fire in a group, manufacturing or mending their tools and engaged in conversation. Mithen thinks that the evidence in favour of the Neanderthals having language is compelling, but concludes that the use of language must have been restricted to the social domain. No one would have given instruction on how to make tools. Above all, the Neanderthals had only a "rolling, fleeting ephemeral consciousness about their own knowledge". There was no introspection — and no art or ritual.

In the modern mind, all the kinds of intelligence are connected. The early human mind could interpret natural symbols such as hoofprints, communicate intentionally in social matters and produce artefacts. But all these were isolated abilities, and it took the modern human mind to be able to "create artifacts and images with symbolic meanings as a means of com-

munication — i.e. art". Anthropomorphism and totemism are the cross-product of thinking about animals and thinking about humans — again, only possible in a mind where natural-history and social intelligence can communicate. Racism, thinking about people as objects to be manipulated, is the product of thinking about objects to be manipulated (technical intelligence) and thinking about people.

Finally, agriculture arises from four aspects of the change in the structure of the mind: specialized tools from the interaction of technical and natural-history intelligence; use of animals and plants from the integration of social and natural-history intelligence; domestication of plants and animals from integration of social and natural-history intelligences; and the manipulation of plants and animals from the combination of technical and natural-history intelligences.

It is a pity that Mithen made little attempt to put his ideas in relation to those in *Origins of the Modern Mind* by Merlin Donald (Harvard University Press, 1991), which, as the title indicates, has a similar objective and uses the same methods of integrating data and ideas from psychology and archaeology. Comment is relegated to a footnote, where Mithen, to give him credit, does "highly recommend" that readers also read Donald. Mithen comments that the complexity and variability of the



VULCAN at his forge teaches humans the arts that distinguish them from animals. This painting by Piero di Cosimo, illustrating a Renaissance theme in anthropology, is one of many pictures in *The Discovery of the Past: The Origins of Archaeology* by the French scholar Alain Schnapp. The book interweaves history, art and literature to trace how Europeans, from antiquity to the age of Darwin, explored their past. British Museum Press, £25.

the archaeological data are insufficiently appreciated in Donald's book. The subtitle of Mithen's book — "A Search for the Origins of Art, Religion and Science" — is a clear come-on for the general reader, but Mithen stops well short. Donald, on the other hand, who goes on to consider literacy and its effects, and the use of artefacts as memory aids, goes much further down this line.

Of course, as a psychologist, I believe all the archaeology Mithen lays before us, but the psychology remains that of an enthusiast. His notion of acceptable evidence will strike most psychologists as odd. He quotes Oliver Sacks on his impression of an autistic whose "sense of animals' moods and feelings is so strong that these almost take possession of her, overwhelm her at times".

This is taken as "good evidence that the mind has a specialized device for learning about the natural world". Here, as in other examples, Mithen fails to distinguish reliably between genetic evolution and cultural evolution, between specialized cortical devices and the consequences of massed learning of a topic. He also seems unaware of the need to make clear distinctions between specialist skills and the sometimes very abstract specialist cognitive abilities that underlie these skills.

In the end, though, this is a very good read, packed with well-presented information and provocation. □

John Morton is at the MRC Cognitive Development Unit, 4 Taverton Street, London WC1H 0BT, UK.

Sport on the brain leads to pain

John C. Marshall

Why Michael Couldn't Hit, and Other Tales of the Neurology of Sports. By Harold L. Klawans. W. H. Freeman: 1996. Pp. 308. \$22.95.

AN intimate relationship between sport and cognitive neurology has existed since the inception of the discipline. When the 26-year-old Baron Edouard de Rampeau was injured in a fencing accident in 1807, he was referred to the eminent physician Franz Joseph Gall (1758–1828), and a discovery about the brain ensued.

The broken tip of the fencing foil had penetrated to the posterior part of the anterior left lobe of the brain. De Rampeau continued to recognize objects and use them appropriately but, Gall reported, "the memory of names was wholly extinguished". The left frontal region therefore came to be seen as crucial to the exercise of the language faculty.

In Gall's day, limited brain damage in

previously healthy young adults was commonly due to the foil, whether used in sport, jest or earnest. Gall eagerly seized on such cases, regarding them as better suited for showing localized functions of the brain than the more extensive effects of apoplexy (stroke) in the elderly.

Fencing does not figure in Harold Klawans' latest collection of neurological essays for the layman. But boxing does. For the most part, boxing, the only 'sport' whose purpose is to cause chronic progressive brain damage, produces diffuse neuronal injury leading to dementia. In some instances, however, the lesions can be more restricted, although their effects on action, cognition and personality are not necessarily less severe. Klawans' chapter on Muhammad Ali is both a fine tribute to 'The Greatest' and a lucid introduction to the functions of the substantia nigra, damage to which was responsible for Ali's post-traumatic Parkinsonism. The consequences of the speed of Ali's lip were,

if anything, even more disastrous than the sad shuffle of his later years.

Klawans is somewhat ambiguous in his own attitude to boxing. He manages to regard it as "one of our eternal verities", while reminding us that two millennia ago the Emperor Augustus "put an end to boxing because the sport was ruining too many prospective soldiers".

Klawans is also strangely unsympathetic to Primo Carnera, the only native Italian to win the world heavyweight championship. Medically, Klawans is excellent on how Carnera's gigantism (acromegaly) was caused by a tumour of the pituitary gland. But he fails to appreciate Carnera's outstanding intelligence: most of Carnera's fights were won by his arranging for the Mafia to pay his opponents to fall over, and he ended his career playing bit-parts in Hollywood movies.

It is a pity that such sensible career-planning is so little stressed in these essays. Top athletes presumably require superior strategic intelligence at least as much as they need strength and visuo-motor coordination. The issue arises only in the chapter on Roger Bannister, where it is emphasized that the four-minute mile was achieved by serious thought and a profound knowledge of physiology rather than by intrinsic talent and intensive application.

Even in these days of supercharged professionals, the thought of Bannister training for half an hour a day "during his lunch breaks from his duties at the hospital" can still bring a tear of cultural joy to the amateur English eye.

For the rest, the sport in Klawans' book is resolutely American, although the diseases are international: baseball and occlusion of the right subclavian artery when pitching, volleyball and Marfan's syndrome (a disorder of connective tissue), basketball and tics (Gilles de la Tourette's syndrome), American football and myasthenia gravis (a disease of the neuromuscular junction), golf and the dystonias (abnormal, involuntary movements that "move part of a limb into an untoward position and keep it there")... and more and more and more baseball injuries.

Klawans' ability to describe complex medical issues in understandable language is peerless. But the jargon of baseball defeated me to much the same extent that cricket cannot be explained to Americans. I did, nonetheless, come away with the strong feeling that many modern sports (and baseball in particular) are extremely dangerous.

Perhaps we should revert to duelling, a comparatively safe sport and one whose underlying code of honour might enforce higher moral standards in public life. □

John C. Marshall is in the Department of Clinical Neurology, University of Oxford, Radcliffe Infirmary, Woodstock Road, Oxford OX2 6HE, UK.

IMAGE
UNAMAGABLE
FOR A CORPUSCULE
FOR A CORPUSCULE
REASONS

Tragedy in the making: Ali beats Sonny Liston (left) but later succumbs to brain damage.

The specific notion

Fred S. Rosen

Exquisite Specificity: The Monoclonal Antibody Revolution. By Alberto Cambrosio and Peter Keating. *Oxford University Press*: 1996. Pp. 243. £39.50, \$59.95.

EVER since Dr Charles McIntyre collected a urine specimen from a greengrocer with myeloma in Devonshire Street, London, in 1846, the study of the disease has been a productive route to discovery. McIntyre, thinking that the greengrocer's urine would contain protein ("animal matter") because he was losing weight rapidly, heated the urine and found that a precipitate formed between 45 °C and 60 °C, which redissolved when the urine was boiled. He then sent this urine specimen to Professor Henry Bence-Jones, at Guy's Hospital, asking "What is this?". Bence-Jones never answered, but the precipitate nevertheless became known as Bence-Jones protein. It was only the second protein to be obtained in pure form (the first having been hen egg albumin). It was more than a century later that Bence-Jones proteins were identified as the light chains of immunoglobulins.

Myelomas are tumours of plasma cells, usually in the bone marrow. They are found in many species, and can be induced in mice by intraperitoneal injection of mineral oil. Myeloma proteins, which are immunoglobulins secreted by malignant plasma cells, eventually helped to solve problems of immunoglobulin genetics and structure. Gerald Edelman shared the Nobel Prize for Physiology or Medicine in 1972 after completing the sequence of a myeloma protein, Eu, the first immuno-globulin to be sequenced completely.

In 1975, a landmark paper was published by Georges Köhler and César Milstein (*Nature* **256**, 495; 1975). These investigators fused mouse myeloma cells with splenic lymphocytes from a mouse that had been immunized with sheep red blood cells. The immortalized hybrid cells now produced prodigious quantities of pure monoclonal antibodies to sheep red blood cells. Standardized reagents, hitherto unknown in immunology, were now available. Several immunologists in the early 1970s were trying to establish monoclonal antibody production by transformed cells, but Köhler and Milstein were the first to succeed.

They grasped the power of this new tool, but could not anticipate that it would spawn a multimillion-dollar biotechnology industry, which came about largely

because of the almost simultaneous development of the fluorescence-activated cell sorter. Monoclonal antibodies worth \$574 million are said to have been sold in the United States alone in 1992. This transformation of a simple and beautiful discovery into a vast biotechnology industry is now recounted by Alberto Cambrosio and Peter Keating in *Exquisite Specificity*. It is in many respects a sordid tale of specious claims to priority, lawsuits and counter-suits, and distortions of science funding that do little credit to the scientific community, and even less to the biotechnology industry. The story is well told in this book.

If Köhler and Milstein had patented their finding, a great deal of grief might

IMAGE
UNAMIGABLE
FOR A COPYRIGHT
FOR REASON
REASONS

Culture club: 'sordid' tale of biotechnology uncovered.

have been avoided. It seems that the British patent office or the Medical Research Council turned down the idea because they did not find the discovery to be meritorious in this regard. Anonymous administrators responsible for such decisions should be publicly exposed for their bad judgement and incompetence. Perhaps the time has come to restore the stockades and gallows at Tyburn as a way of reintroducing accountability.

This history book should be of interest to the whole biology community. It is well illustrated to guide those not familiar with hybridoma technology through its details. Unfortunately the prose sometimes gets a little turgid. For example, try the sentence: "Explicitly or implicitly, to speak of a context is to reduce whatever is being contextualized to the elements that make up the context". In the foreword, Milstein quotes Jorge Luis Borges: "sólo quedan palabras. Palabras, palabras desplazadas y mutiladas" ("there remain only words. Words, displaced and mutilated words").

No exegesis on one of the most important discoveries of our century can diminish the lustre and impact of the paper by Köhler and Milstein. They were rewarded with the Nobel Prize for Physiology or Medicine in 1984. □

Fred S. Rosen is at the Center for Blood Research, 800 Huntington Avenue, Boston, Massachusetts 02115, USA.

Tunnel vision

I. Bernard Weinstein

Racing to the Beginning of the Road: The Search for the Origins of Cancer. By Robert A. Weinberg. *Harmony*: 1996. Pp. 271. \$27.50.

DURING the past few decades, our understanding of the origins of cancer has been tremendously enriched by the discovery of cellular oncogenes and tumour-suppressor genes. Robert Weinberg's book provides a nontechnical and exhilarating description of how this field evolved. The format is a series of anecdotes and stories laced with pungent descriptions of individual scientists, as well as examples of collaboration, bitter competition and a few cases of deliberate deception. These events are deftly woven together to illustrate the unpredictable nature of the scientific endeavour and how disparate areas of research can unexpectedly converge to provide new insights and models. Some of these stories are tedious, especially the lengthy descriptions of the research in Weinberg's own laboratory. But he did make important discoveries; this is his book, and he is not a modest man.

It is not apparent why, in glorifying the contributions of molecular biology to cancer research, the author finds it necessary to bash other disciplines, especially as the title of the book promises to take us "down the road", and not just one of the paths, in the search for the origins of cancer. For example, in an early chapter entitled "A can of worms" he states that by the 1960s the field of chemical carcinogenesis had reached an impasse and researchers in this field "became increasingly demoralized, seeing no clear way to advance their work". But for those of us who were there, this field was and still remains an exciting area of cancer research — and not because we are on Prozac.

Major discoveries were made, including the principles of metabolic activation and detoxification of carcinogens, the structures and mutagenicity of specific carcinogen-DNA adducts, the role of DNA repair as a host defence mechanism and basic principles of tumour promotion and hormonal carcinogenesis. These findings also provide profound insights into the origin of cancer, including the origin of specific mutations in *ras* oncogenes and the *p53* tumour-suppressor gene.

Weinberg glosses over these accomplishments by numerous investigators in the United Kingdom, the United States, Japan and elsewhere, and instead states that Bruce Ames's aphorism in 1979 that carcinogens are mutagens "had become the credo of our religion", even though in recent years Ames has backed away from this equation, recognizing that it is an over-

James Holmes/SPL



In his enchantingly illustrated children's book *Starry Messenger*, Peter Sis, an internationally acclaimed illustrator and film maker, tells the story of the life of Galileo. Its publication today comes exactly four years after the Vatican's rehabilitation of the grand old man. Farrar, Straus and Giroux, \$16.

simplification. Indeed, Weinberg's initial discovery of mutated cellular oncogenes was made possible by the use of well-characterized carcinogen-transformed cells developed by investigators in the field of chemical carcinogenesis.

Also, in an early chapter, he asserts that research on viruses and cancer "threatened to divert a rapid forward thrust straight into another patch of very deep mud." And yet, later in his book, Weinberg recognizes that research on tumour viruses has had an enormous pay-off, including the discovery of reverse transcriptase, the roles of hepatitis B and C viruses in liver cancer, of Epstein-Barr virus in Burkitt's lymphoma, of human papilloma virus in cervical cancer, of a retrovirus (HTLV-1) in one form of lymphoma, and the viral origin of AIDS and AIDS-related malignancies. Not a bad record for a field mired in the mud. Furthermore, some of us would not be surprised to learn that viruses or other infectious agents have a role as cofactors in the origin of other types of cancer.

Weinberg is also merciless in his criticisms of those who have emphasized epigenetic mechanisms in carcinogenesis, even though there is substantial evidence that both epigenetic abnormalities and mutagenesis play important parts in the multistep process of carcinogenesis. He

also glosses over the recent advances made in cancer immunology, cancer biology (apoptosis, angiogenesis, invasion and metastasis) and epidemiology. He does, however, acknowledge the importance of clinical and epidemiological studies in revealing cigarette smoking (a source of chemical carcinogens) as the cause of lung cancer. Perhaps this is because he claims Ernst Wynder, but not of course Sir Richard Doll, as a distant relative.

In the euphoria surrounding recent advances in the molecular biology of cancer, we must remain sober about the limitations in translating our knowledge and accomplishments into more effective strategies for cancer prevention and treatment. We still lack hard facts about the exogenous factors that cause cancers of the breast, prostate and colon, even though today they account for about half of new cancer cases in the United States and Western Europe and their incidence is increasing in other countries. Although dietary factors have been implicated, experts in the field disagree on the precise components and the underlying mechanisms.

Within the mechanistic framework of oncogenes and tumour-suppressor genes, the multistep evolution of cancer cells seems to be much more complex than was

originally thought when the first oncogenes were discovered. We now know that a given tumour often carries several mutated genes as well as gross chromosomal abnormalities, that more than a hundred different genes have been found to be mutated and/or abnormally expressed in various types of human cancer and that each year this list gets longer. The biochemical functions of these genes are extremely diverse, and encompass intracellular circuitry involved in signal transduction, gene transcription, cell-cycle control, DNA replication and repair, differentiation, senescence and apoptosis, as well as extracellular functions related to cell-cell and cell-matrix interactions, proteases, locomotion and angiogenesis.

These complexities do not of course detract from the recent progress properly celebrated in Weinberg's book. They do, however, indicate the need for greater modesty and for the development of more holistic and multidisciplinary approaches in cancer research that will complement the reductionist approaches emphasized in the past few decades. □

I. Bernard Weinstein is at the College of Physicians and Surgeons, Columbia University, 701 West 168th Street, New York, New York 10032, USA.

Rocket blast from the past

Michael J. Neufeld

The Peenemünde Wind Tunnels: A Memoir. By Peter P. Wegener. Yale University Press: 1996. Pp. 187. \$30, £19.95.

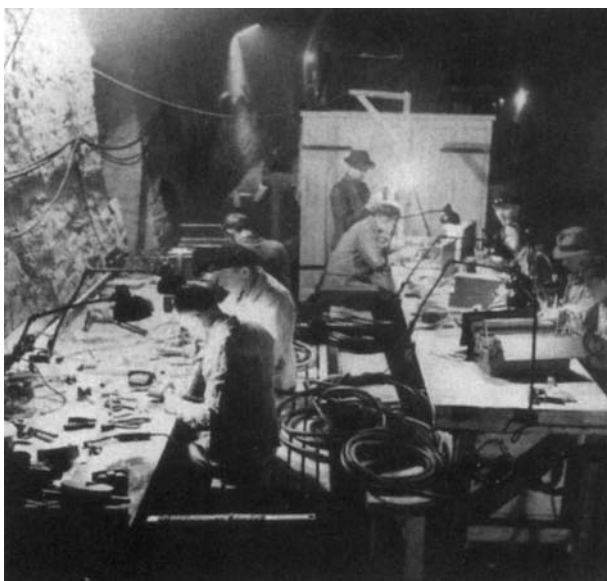
PEENEMÜNDE: the name of the Third Reich's leading rocket centre had and has a certain magic. Peter Wegener spent less than a year there, in 1943–44, as well as another year or so working with the wind tunnels after they were evacuated to the Bavarian Alps. Yet that year or two were formative in his life; they turned the aspiring geophysicist and veteran of the Eastern Front into an aerodynamicist. They also exposed him to Wernher von Braun — the charismatic technical director of that revolutionary enterprise — and to the opportunities and dilemmas of working on military research for a totalitarian regime.

Despite being sincerely anti-Nazi, Wegener came to grapple with the moral dilemmas only over time, and to some extent only recently. After decades of obfuscation by von Braun and his followers, the name Peenemünde — which had come to signify for many the first step on the road to space — finally became wedded in the 1980s to another name: Dora, the concentration camp that supplied the slave labour for V2 missile production. The efforts of survivors of that camp, and the forced departure from the United States of Arthur Rudolph who had been production director in the underground plant, made it difficult to think of one without the other. Yet angry denial of Rudolph's guilt, and of any responsibility for the prisoners' fate, has remained the nearly universal reaction of the former Peenemünders. Wegener's genuine attempt to deal with the implications of Dora therefore makes this absorbing memoir a milestone in the literature about the centre.

Like most of the scientific and engineering staff there, Wegener was drafted into Peenemünde knowing little or nothing about rocketry. The son of a well-known Berlin actor and film director and nephew of Alfred Wegener, the great geophysicist and advocate of 'continental drift', Peter Wegener came from a liberal bourgeois family with a natural aversion to Nazism. Born in 1917, he reached his maturity at the apogee of the Third Reich and saw little choice but to live with the system while circumventing its ideological demands as much as possible. His academic training was repeatedly interrupted by service in the Luftwaffe anti-aircraft artillery, most notably in Russia, but he

finished an abbreviated doctorate in 1943 and was almost immediately sent to the air-force group at Peenemünde which was working with the army on anti-aircraft missiles. He found himself assigned to Rudolph Hermanns institute where groundbreaking work was being done in supersonic aerodynamics.

Those who come to this short memoir expecting detailed discussion of Wegener's scientific work will be disappointed. But Wegener does give valuable



V2 rocket builders: Germany's wartime underground plant.

new insights into the aerodynamics institute, its personalities and its director, who comes off as a rather unsavoury believer in Hitler's regime. Wegener also describes his transfer to the United States in 1946 under Project Paperclip, working with a small group to reconstruct the Peenemünde tunnels at the Naval Ordnance Laboratory outside Washington DC. He eventually became a distinguished professor at Yale. For many readers, however, the most engaging parts of the book will be those where he discusses his encounters with von Braun, the underground plant and the realities of working under that regime.

His answers will please neither of the extreme camps that dominate opinion about von Braun and Peenemünde: the traditional hero-worshippers and those who see only goose-stepping Nazis. On the one hand, Wegener has only good things to say about von Braun, whom he describes as "gifted and complex", "one of the great engineers of our time" and "cat-

egorically... no Nazi". "His outlook on life," Wegener asserts, "was far removed from the prevailing ideology." On the other hand, Wegener attacks the "myth" that Peenemünde was a space centre and that its military work was just a "detour" on the road to space. He concludes that Rudolph was a "Nazi" and was guilty of initiating the exploitation of concentration-camp labour at Peenemünde before a Royal Air Force raid forced production underground. Wegener's single two-hour visit to the underground plant, Mittelwerk, in about March 1945 left a lasting impression: "I have never before experienced such glances of hate" — from the prisoners. He questions how those assigned from Peenemünde could have kept their sanity. He also mentions von Braun's many visits there, but accepts

Ernst Stuhlinger's assertion that von Braun struggled with his conscience over what to do about the prisoners.

This last point will irritate von Braun's critics. The truth is that we do not know — we will never know — what he thought about forced labour at the time, because no written evidence exists. Stuhlinger's story comes from conversations with von Braun decades later, after the French survivors of Dora had forced von Braun to rationalize his role. But we do have written evidence of his involvement in decision-making about concentration-camp labour from at least August 1943 on. As to his party and SS memberships, I could accept Wegener's assumption that these were reluctantly accepted political necessities for von Braun, who was a nationalist but no Nazi ideologue. Yet, what-

ever von Braun's beliefs, his actions again put him in a less flattering light. Wegener comes nearest the truth early in the book when he says that von Braun was "close to being obsessed with rocket development" and "disliked Hitler and all Hitler did [a statement truer after 1943 than before]. But Hitler supported his dream."

Whether one accepts all of Wegener's judgements or not, in the last analysis this memoir of Peenemünde is unique because it breaks the traditional mould. His account of the wind tunnels is also informative, interesting and useful to historians. Despite instances of awkward writing that cry out for a stronger editorial hand, the book is also readable. It belongs on the reading list of all those interested in the history of rocketry, as well as anyone fascinated by the dilemmas of science and technology in the Third Reich. □

Michael J. Neufeld is at the National Air and Space Museum, Smithsonian Institution, Washington DC 20560, USA.

To directly go

Christopher P. McKay

The Case for Mars: The Plan to Settle the Red Planet and Why We Must. By Robert Zubrin with Richard Wagner. Free Press: 1996. Pp. 318. \$25, £16.99.

"THE Case for Mars" is the title of a series of conferences held at the University of Colorado every three years since 1980. The goal of the conferences has been to develop the case for human exploration of Mars and to devise practical methods of achieving this goal. This book, named after these conferences, is a concise statement of the state of the art in mission designs.

The most important difference between the scenarios for Mars exploration developed in *The Case for Mars* and earlier visions such as those of Wernher von Braun in the 1950s and the short-lived post-Apollo studies is the reliance on resources produced on Mars from Martian materials. This bold live-off-the-land

philosophy has enabled Robert Zubrin and his fellow rocket scientists to devise a mission that would send a crew of four to Mars in 2007 for a total cost of \$20 billion. This is a small fraction of what the conventional wisdom deemed to be the cost of human exploration of Mars. How does Zubrin do it? By going directly to Mars and forgoing all the way-points and research programmes that many argue are prerequisites on the path to Mars. His plan, appropriately called "Mars Direct", has no use for a space station — international or otherwise — nor for an expensive decade-absorbing lunar base. Go to Mars, go directly to Mars, do not....

This plan is faster, better and cheaper because of the small crew, artificial gravity provided by rotation, acceptance of the cosmic radiation accumulated during transit and a safe-haven strategy that assumes that humans can survive on Mars while they await rescue.

Where using Martian resources really generates a payback is on the surface of Mars itself. In contrast to the limited exploration of the Moon possible with Apollo,

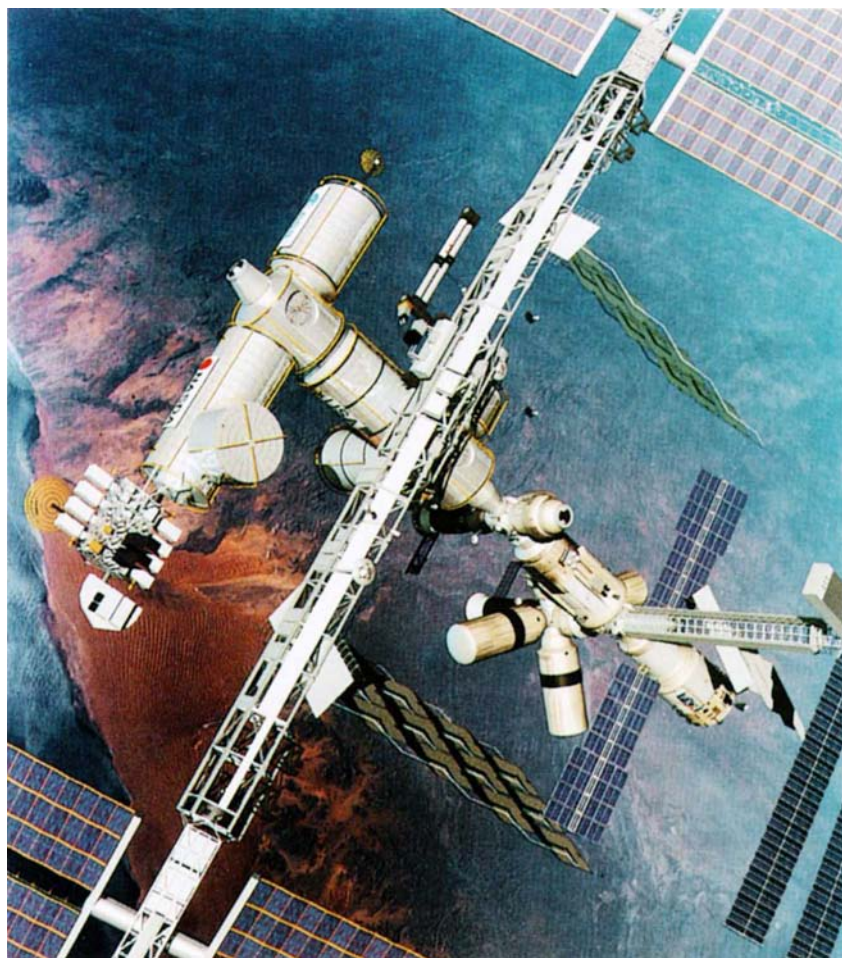
The Case for Mars sees wide-ranging exploration of the canyons, craters and ancient river features of Mars. This global reach of exploration is required if humans are to unravel the secrets of past life on Mars. Drawing on many diverse fields and published works, Zubrin makes the case that life on Mars — past, present and future — is the answer to the question "Why Mars?". Did life arise on Mars during its past water-rich epoch and was this life similar, even genetically related, to Earth life? If so, does this life persist now in a subsurface niche? And in either case could life be introduced to Mars again in the future? No other scientific goal would justify human exploration or requires it the way a life-search does.

These future explorers ("Raiders of the Lost Biosphere") would breathe oxygen made from Martian carbon dioxide, drink water wrung from the Martian atmosphere and drive vehicles with an engine powered by methane and oxygen — both produced on Mars — with about as much horsepower as the average car. The roads won't be good but the scenery will be spectacular and the traffic light.

From the initial human exploration of Mars, Zubrin turns to longer-term projects of making Mars Earth-like: terraforming. Here is the ultimate use of Martian resources for life support. Not just the support of an isolated outpost of humanity, but support of a planetary-scale biosphere. Zubrin offers the reader a short course in the mathematics of warming planets. A few might find the equations and graphs hard to digest, but the meal is worth the effort. A combination of super-greenhouse gases (such as we are now ill-advisedly releasing into the Earth's atmosphere) and orbiting solar mirrors warms Mars enough for its own inventory of carbon dioxide to evaporate from the polar caps and desorb from the soil. The positive feedback that results will warm Mars in timescales measured in centuries. There is no lack of vision here.

Zubrin is unassailable on technical subjects but he mars his book by turning at the end to finance and sociology. Can we really believe that a prize is the way to fund this great human adventure? And can any serious student of history think that white settlers expropriating Indian lands in the American West is a useful metaphor for Mars exploration? No, the exploration of the sea and scientific outposts in Antarctica are the examples that are relevant here. Despite this lapse, *The Case for Mars* is a well-considered blueprint for human exploration of the 'red planet', and I recommend it for anyone who would like to see humans on Mars in our lifetime. □

Christopher P. McKay is in the Space Science Division 245-3, NASA Ames Research Center, Moffett Field, California 94035, USA.



In *Inland in the Sky: Building the International Space Station*, Piers Bizony reports on the challenges of the most ambitious space project since the first Moon landings. He investigates why the project took 12 years to get started, and describes earlier space adventures from Salyut to Skylab. Aurum, £14.95 (pbk).

Environmentally friendly

Walt Patterson

Promising the Earth. By Robert Lamb. Routledge: 1996. £30 (hbk); £9.99 (pbk).

THE environment is no longer the issue it was 25 years ago. In the early 1970s, the environment was front-page news. In the run-up to the UN Conference on the Human Environment in Stockholm in 1972, print and broadcast media were saturated with coverage about pollution, waste and other assaults on the planet. 'Ecology' became not a science but a bar-room cliché.

In the 1990s, by contrast, the media tend to feature the 'contrarians', who dismiss all talk of global warming or resource depletion as alarmist nonsense, and disparage organizations such as Friends of the Earth (with its contradictory acronym FoE) and Greenpeace as at best mischievous, at worst 'eco-fascists' and 'eco-terrorists'. The paradox is that the 'contrarians' now attract media attention precisely because their argument is so out of step with accepted opinion. In the past 25 years, the environment as an issue has evolved from a marginal if irksome afterthought into an essential dimension of international, governmental and corporate policy almost worldwide.

How did this extraordinary metamorphosis come about? Robert Lamb's book offers some enlightenment. One-time head of information services for the United Nations Environment Programme, Lamb is an experienced and knowledgeable commentator. Commissioned by FoE in Britain to write a history of its first 25 years, he has carried out a potentially difficult task with commendable flair. Throughout its quarter-century, FoE has engaged the dedicated commitment of many able, articulate and combative individuals, working at full stretch, often against daunting odds. Conveying accurately the flavour of such activities, including the inevitable internal tensions, is not easy. I know: in the 1970s I was the first energy campaigner for FoE UK, and my short history of the first decade appeared in 1984 as a chapter in a book by a former FoE chairman, Des Wilson.

Despite some minor niggles — misspelled names and some inaccurate chronology — Lamb weaves an engrossing narrative. Studded with sharply etched personal portraits and vignettes, it captures vividly the feel of those early days, when the environment excited radical passions and provoked furious controversy, with environmental activists on one

IMAGE
UNAVAILABLE
IMAGE
FOR COPYRIGHT
UNAVAILABLE
REASONS
FOR COPYRIGHT
REASONS

Fashion statement: environmental campaigning has come a long way since the 1970s.

side and governments and corporations almost invariably on the other.

But the book is more than a history of a single organization. It is a thoughtful overview of the development and evolution of environmental concern and campaigning for more than a century, focusing on FoE UK but setting its activities in a much wider context.

Lamb identifies and scrutinizes key themes that still echo through environmental issues in the 1990s. Can rational scientific and economic argument win a policy decision against the opposition of powerful government and corporate interests? Or must activists take to the streets and the barricades to defend the environment against the destroyers? Lamb analyses the Windscale (later Sellafield) nuclear reprocessing controversy of the 1970s as an example of the former approach, and the campaign to stop the Newbury bypass road in the 1990s as an example of the latter.

He also contrasts FoE's campaign tactics with those of Greenpeace as intensively researched lobbying versus media spectacles. The contrast is overstated. Even in the 1970s, the most effective FoE campaigns combined solid research and policy development with the multiplicative potency of effective media relations and street-level popular support. A more striking contrast, perhaps, is that between the centrally directed global operations of Greenpeace International and the decentralized network of more than 50 autonomous national member organizations of Friends of the Earth International, one of the oldest being FoE UK.

In the 1990s, however, environmental campaigning is no longer primarily about sounding the alarm. Despite the protests of the contrarians, the urgency of environmental concern has been broadly acknowledged at the highest government

and corporate levels. The UN Conference on Environment and Development in Rio de Janeiro in 1992 initiated a global process, whose progress, however uneven, continues to gather momentum.

As we approach the twenty-first century, environmental policy will focus more and more on striving for sustainable solutions. As Lamb describes, the issues are less polarized than in the 1970s, partly because alumni of FoE and other activists are now in senior political and corporate positions. Environmental campaigning is not what it was. It may be better. It had better be. □

Walt Patterson is at the Royal Institute of International Affairs, Chatham House, 10 St James's Square, London SW1Y 4LE, UK.

Femme fatale?

Roy Porter

Typhoid Mary: Captive to the Public's Health. By Judith Walzer Leavitt. Beacon: 1996. Pp. 331. \$25.

TYPHOID Mary pops up time and again as one of the colourful villains of medical history — indeed a *femme fatale* in the most literal sense. She was the cook who, while healthy herself, infected nearly 50 people with enteric fever in New York early this century. Her tale forms a *cause célèbre* because it contains all the gripping elements of a good detective story.

From 1900, mysterious typhoid cases multiply in swish New York households; medical investigators are called in; and a bacteriological sleuth points the finger of blame at an Irish-born cook called Mary Mallon. She struggles with a carving fork

but is apprehended and mandatorily detained in a hut in the grounds of an isolation hospital on North Brother Island, on New York's East River.

Some years later, after a *habeas corpus* suit, the health department finally permitted her release, on condition that she worked in a laundry; but she reneged on her agreement, reverted to cooking and, under an assumed name, got a job in a maternity hospital kitchen ("why not?" she thought — after all, she'd never been sick with typhoid herself). The result? A fresh outbreak, with two more people dying; rearrested, she was sent back to Quarantine Island and this time detained there, alone with her fox terrier, for life — another 23 years — banished "like a leper", as she put it, and pleading her innocence to the end.

What was her crime? Her gall-bladder was teeming with the rod-shaped typhoid bacillus. Little was then understood about being a carrier. It was only recently that the phenomenon itself had even been discovered, thanks to the new bacteriological advances due to Louis Pasteur and especially to Robert Koch. Koch's already famous 'postulates' had shown the way in which distinctive infectious diseases could be ascribed to a particular microscopic causal agent. But his exemplary laboratory techniques also brought to light something very peculiar: the fact that some people had the alarming capacity to harbour such contagious bacilli with no harm to themselves — that they were, so to speak, immune.

The bacteriological theory behind the carrier remained little understood, but the New York health authorities had little doubt what action they needed to take. The United States believed in personal liberty no doubt, but it also believed in public health; and here was a case in which the former must yield to the latter. And because Mary Mallon was the first carrier apprehended in the United States, her treatment was designed to set an example.

In taking her fresh look at Typhoid Mary, the distinguished medical historian and feminist scholar Judith Leavitt goes beyond the hackneyed yarn, punctures some myths and sets the story in its full historical context. She highlights how the case is regularly served up as a triumph of the new bacteriology, forgetting the woeful fate of Mary herself. She was rapidly demonized: by 1909, the press had invented the Typhoid Mary label, called her a "fiend" and produced cartoons of her cracking egg-like skulls into a frying-pan. Through such images, Mary Mallon foreshadows all those other *femmes fatales* so familiar in US pulp culture — one thinks of misogynistic movies such as *Fatal Attraction* and *The Hand that Rocks the Cradle* in which insidious anti-heroines worm their way into the

The personalities who have devoted their lives to studying and treating tropical diseases figure strongly in *The Wellcome Trust Illustrated History of Tropical Diseases* edited by F. E. G. Cox. Tropical disease was first recognized as a separate branch of medicine at the turn of the century but the book traces the history of the subject back to the earliest written records. The angel of death dates back to a yet earlier age; this painting of the spectre dropping deadly substances into a river is meant to symbolize typhoid. The book is aimed at a general scientific audience and examines the status and treatment of individual diseases. Wellcome Trust, £35.



homes of respectable America and then, witch-like, callously fill them with *maleficium*, poisoning family relations and even the dessert.

Taking the broader perspective, Leavitt asks what the Mary Mallon story tells us about public health in the new age of bacteriology. She draws attention to the barely concealed prejudices directed against Mary Mallon as an Irish immigrant. Some 25 million migrants entered the United States between 1880 and 1924, and once they were there certain groups became targets of vituperation: the Chinese were blamed for plague in San Francisco in 1900, the Italians for polio in New York, and the Irish were widely condemned as drunk and dirty.

Then, as is signalled in her book's title, questions are raised about the powers of public health. Mary Mallon ended up permanently institutionalized, not because she had committed a crime but because she was deemed a threat to health. Such use of preventive detention on medical grounds had become entrenched by the dawn of the twentieth century: witness the building of tuberculosis sanatoria as, in effect, jails for the sick poor.

In Victorian England, the Contagious Diseases Acts had already led to the locking up of thousands of alleged prostitutes suspected of having venereal disease. And more than 20,000 prostitutes were soon to be quarantined in the United States during the First World War, so as

to ensure the health of the troops. In Illinois, the Board of Health assumed the authority to hospitalize any woman thought to be venereally infected. If she refused, the board posted a large placard on her home reading "Suspected VD". "To drain a red-light district and destroy thereby a breeding place of syphilis and gonorrhea," one federal official explained, "is as logical as it is to drain a swamp and thereby a breeding place of malaria and yellow fever." It is no accident, Leavitt suggests, that it was mainly females who were in this way vulnerable to victim-blaming as threats to public well-being.

One of the themes of this alert and thoughtful work is the endless return of the ghost of Mary Mallon in the media and popular memory. It is after all a story that continues to have relevance, especially in the age of AIDS and of drug-resistant tuberculosis. How do we balance public salubrity against sacred individual liberties? Without pretending to offer any simple solutions to the dilemmas posed by modern epidemics, Leavitt counsels us, through her sympathetic re-creation of the tragedy of Mary Mallon, that such decisions can never be cut-and-dried, and should not be seen as narrowly medical. □

Roy Porter is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BE, UK.

Madness on trial

Keri K. Gould

The Power to Harm: Mind, Medicine and Murder on Trial. By John Cornwell. Viking: 1996. Pp. 321. £18, \$24.95.

"BEHIND every crooked thought there lies a crooked molecule." This provocative soundbite is from testimony in the *Fentress v. Eli Lilly and Company* case. John Cornwell, a British journalist, uses the case to explain his concern with the way America's legal system and society have embraced scientific reductionist philosophy and psychopharmacological treatment to resolve individual and social dilemmas. Scientific reductionism seeks to explain human action and emotion by reducing behaviour to the molecular level, isolated from individual will and without social context. The theory removes the community from any responsibility for creating an environment supportive of antisocial behaviour.

Legal drama

Cornwell examines in depth a legal drama that at first glance appears to be about victim compensation for the tragic 1989 mass murder committed by Joseph Wesbecker, a disgruntled printing-plant employee in Louisville, Kentucky. The 12 survivors and 16 relatives of the eight deceased victims filed suit against the pharmaceutical giant Eli Lilly. (The other original defendants were eventually dropped from the case or settled out of court.) Lilly makes Prozac, the antidepressant medication Wesbecker had been taking for a month before he took an AK-47 assault weapon to his former employment site and opened fire, seriously wounding or killing 20 people before killing himself.

The plaintiffs contended that Lilly was legally liable because Prozac had driven Wesbecker to commit his crime. They claimed that Lilly was negligent in making and selling Prozac because the company knew or should have known that the drug can induce violent episodes in patients with no history of violent behaviour. The plaintiffs said that the clinical tests Lilly used to persuade the federal Food and Drug Administration (FDA) that Prozac was safe to market were inadequate and the results carefully choreographed from sophisticated statistical skulduggery. It was claimed that Lilly deliberately failed to inform the FDA about serious problems with the drug. The plaintiffs' complaints also included breach of warranty and strict liability claims which asserted that the drug was unreasonably dangerous and that the company had failed to warn the public of those dangers. Lilly denied every allegation and produced voluminous testimony to defend its scien-

tific trials and reporting practices.

Even when the facts of a case seem deceptively simple, the assignment of causality and blame is far more complex. In Cornwell's view, the jury was not provided with a reasoned basis for deciding liability for this tragedy. Rather, the case became a *cause célèbre* for the vindication of a corporate pharmaceutical giant's cash cow, and a legal battle between free will and neurophilosophical reductionism. It provided the breeding ground for a secret pact and subsequent cover-up agreed to by the parties without the knowledge of the presiding judge.

At the heart of the cover-up controversy were the presiding judge's eviden-

The case became a *cause célèbre* for the vindication of a corporate pharmaceutical giant's cash cow, and a legal battle between free will and neurophilosophical reductionism.

tiary rulings concerning the introduction of damaging testimony on prior criminal sanctions levied against Lilly for failing to inform the FDA of deaths blamed on its drug Oraflex, an anti-arthritis medication sold in Britain before receiving US approval. After the final jury verdict of nine to three for the defendant Lilly, Judge Potter became suspicious when the plaintiffs failed to appeal. He suspected a secret settlement payoff. This was disputed by both parties, and erupted into vitriolic litigation against the judge, who sought to change the final judgment from "dismissed with prejudice" to "dismissed with prejudice as settled".

Cornwell's thorough, wide-ranging investigation of the incident, the individual and corporate entities involved and the political climate make for fascinating reading, although the sheer amount of information contained in the book and the chapter-by-chapter shifting of time, place and narrative voice can be disruptive. Reading a succinct article highlighting the facts of the incident and the procedural trial manoeuvres after I read the book gave me a better framework for appreciating the enormous amount of information assembled by Cornwell.

The book's shifting storyline dilutes the strength of the author's anger that Wesbecker's story and its exposure of the dangers of Prozac were overborne by Lilly's ability to orchestrate an enormous secret money settlement which ensured that the drug's reputation could not be

sullied. The company's chutzpah was able to circumvent honest jury deliberations and promote the potential abuse of the judicial process.

Despite Cornwell's admirable research, several of the issues raised are inadequately or incorrectly explored. First, Cornwell decries the public's acceptance of increasingly intolerable workplace conditions, illustrated by Wesbecker's unrelenting harassment by the management at the printshop. His salary was cut, while it became common for him to work three or more 16-hour shifts a week in dangerous circumstances, despite medical support for his requests not to work "the folder", the most stressful and onerous task.

Facing up to the types of difficulties experienced by Wesbecker, but tragically too late to help him, the Americans with Disabilities Act was signed into law in 1990. Congress had concluded that insidious forms of discrimination against individuals with disabilities, including isolation and segregation, persist in such critical areas as employment, housing, education, health services and access to public services. Before the act, as in the case of Wesbecker, people with disabilities often had no legal recourse to obtain from an employer a reasonable recognition of their disability.

Second, the case raises important questions about the use of scientific evidence in court and the competence of judges and jurors to evaluate such evidence. Cornwell suggests that, in cases where the verdict depends on the understanding of complex scientific knowledge, judges should decide the verdict rather than juries. I think this view is misguided as it assumes judges to be more knowledgeable, less influenced by personal prejudices and better able to understand scientific data than juries.

An examination of the development of mental-disability jurisprudence indicates that judges routinely interject prejudicial viewpoints and pretextual reasoning into the decision-making process. Decision-making based on pretextuality allows judges to accept (either implicitly or explicitly) testimonial dishonesty from witnesses, especially expert witnesses, to achieve a specific end while often subordinating standards of statutory or case law. The amount of adverse media attention showered on the US judiciary, particularly in controversial cases, increases this possibility.

Scientific illiteracy

Additionally, most judges simply do not have the educational background in science to understand any better than some lay jurors the often bewildering explosion of sophisticated genetic and scientific information that lawyers are seeking to bring to court. Recognizing this chasm, the Human Genome Project has funded

the Einstein Institute for Science, Health and the Courts to begin a national programme to educate a thousand state and federal judges in genetics and molecular biology in the next two years and to sponsor conferences in related areas.

Third, the issue of Wesbecker's legal insanity is inadequately explored in the book. Cornwell classifies the insanity defence as an "excuse" defence. He attributes its underpinnings, as well as that of several theories of mitigation, to the *Durham* decision which he characterizes as the rule that acquits individuals of criminal behaviour by reason of "irresistible impulse". Unfortunately, this characterization is blatantly incorrect. In fact, *Durham* specifically rejected both the *M'Naghten* good-evil test and the irresistible-impulse test by finding that a broader test is appropriate.

The *Durham* judgment said: "We find that as an exclusive criterion the right-wrong test is inadequate in that (a) it does not take sufficient account of psychic realities and scientific knowledge, and (b) it is based upon one symptom and so cannot validly be applied in all circumstances. We find that the 'irresistible impulse' test is also inadequate in that it gives no recogni-

tion to mental illness to the application of the inadequate right-wrong test. We conclude that a broader test should be adopted."

Durham predicated the lack of criminal responsibility on a finding that the "unlawful act is the product of mental disease or defect". *Durham* came under rigorous criticism for being too heavily dependent on medical expertise and was overruled in 1972. *US v. Brawner*, which overruled *Durham*, added a volitional component to *M'Naghten*'s cognitive inquiry. The *Brawner* test holds that defendants are not responsible for their criminal conduct if, as a result of mental disease or defect, they "lacked substantial capacity either to appreciate the criminality of their conduct or to conform their conduct to the requirements of law".

Kentucky law offers both a "not guilty by reason of insanity" (NGRI) verdict and a "guilty but mentally ill" (GBMI) verdict. About a third of the jurisdictions in the United States do likewise. Under GBMI, a person may be found guilty of the crime and mentally ill, but not legally insane at

the time of the offence. People convicted under such a statute are sentenced to prison with the caveat that they are mentally ill and should receive treatment during their incarceration. The far more interesting question, not raised in the book, is whether Wesbecker would have been able successfully to use the insanity defence had he lived, and what impact, if any, an NGRI or GBMI verdict would have had on the subsequent civil proceedings.

Cornwell warns of the potential apocalypse created by "the growing crisis over reductionist solutions to individual suffering and social disorder". As society increasingly turns from social solutions to pharmacological quick-fixes, Cornwell foresees dire consequences. But he closes the book with hopeful musings on the enduring power of social solutions and leaves his readers to ponder the imperfect relationships between mind, medicine and murder.

Keri K. Gould is at the School of Law, Fordham University, 140 West 62nd Street, New York, New York 10023, USA.

When litigation becomes a lottery

Lee Loevinger

Science on Trial: The Clash of Medical Science and the Law in the Breast Implant Case. By Marcia Angell. Norton: 1996. Pp. 256. \$27.50, £22. Published in the United Kingdom in January 1997.

MARCIA Angell, executive editor of the *New England Journal of Medicine*, considers herself a feminist and a liberal Democrat but, above all, a scientist. She had no particular interest in silicone breast implants until the journal published two articles on the subject that indicated a discrepancy between the law and the scientific evidence.

By 1992, implants had been on the market for 30 years, between 1 million and 2 million women had them, and more than 90 per cent were pleased with them. But in April that year, David Kessler, commissioner of the US Food and Drug Administration (FDA), announced a virtual ban. At that time there was no scientific evidence that implants caused disease, but women panicked, and lawyers filed more than 30,000 suits against implant manufacturers.

In April 1994 a proposed settlement was approved in a class action that provided US\$4.25 billion to settle all claims. That sum included \$1 billion for lawyers' fees. The settlement provided for payment unconditionally to any woman who had received an implant without any signs of illness or other conditions caused by it. Advertisements inviting women to

file claims appeared throughout the United States in newspapers and on television and radio.

In June 1994 the journal published the first epidemiological study of breast implants, and it found no association with disease. But by May 1995 more than 400,000 women had filed claims, Dow Corning, the principal defendant, was overwhelmed by the number of lawsuits and filed for bankruptcy, and the proposed settlement fell apart. Angell was struck by the discrepancy between the legal proceedings and the scientific evidence, and decided to write this book.

Although hooked on the apparent conflict between science and law, Angell is no zealot. She defends the function of the FDA while questioning, rather than condemning, the implant ban. She reports epidemiological studies done after 1994, and lucidly explains the relevance of the statistics, which continue to show no apparent connection between implants and disease.

What she "was not prepared to find was the extent of the financial involvement" of plaintiffs' lawyers, the complicity of some physicians, and cases resulting in relatively small gains for many women but enormous gains for a few lawyers.

The breast implant story also illustrates how subpoenas can be used to harass and intimidate scientists. The authors of breast implant studies have

ADVERTISEMENTS

Martin Dunitz Publishers



HIGHLIGHTS



Frontiers in Multiple Sclerosis

Abramsky & Ovadia
November 1996 •
ISBN 1-85317-384-3 • £65.00

An Atlas of Malignant Haematology

Mufti, Fladrin, Schaefer, Sandberg & Kanfer
1996 • ISBN 1-85317-054-2 • £115.00

Cytokines and Molecular Therapy

Editors: Talpaz & Herrmann
Quarterly journal as reviewed in Nature
Institutional: UK & EC - £135.00;
USA & Canada \$225.00; ROW £145.00
Personal: UK & ROW - £65.00
USA: \$99.00

Martin Dunitz Publishers has extensive current and forthcoming lists in the following areas:



HAEMATOLOGY



ONCOLOGY



NEUROLOGY

**MARTIN
DUNITZ
PUBLISHING**

The Livery House
7-9 Pratt Street
London NW1 0AE
Tel: 0171 482 2202
Fax: 0171 482 7088

*Scientific publisher of
medical books and journals*

been served with excessive and burdensome document demands from plaintiffs' lawyers that would breach the confidentiality of medical records. These discourage, and may prevent, further scientific research, which will serve the interests of plaintiffs' lawyers but of no one else.

Other countries responded to unfolding evidence about breast implants quite differently from FDA. The United Kingdom imposed no ban, stating that there was no scientific evidence of increased risk of disease from breast implants. The practices causing the plague of lawsuits in the United States do not exist in the United Kingdom.

In such US cases, plaintiffs' lawyers have had contingency arrangements to receive fees of 30 to 40 per cent of all money recovered. Such arrangements are not permitted elsewhere. The large number of US claimants has been due to lawyers soliciting for clients through widespread advertising. British barristers are forbidden to advertise or to solicit clients. Finally, the exorbitant sums awarded in implant cases have been the product of jury sympathy rather than evidence, but only in the United States are such damages decided by juries.

Angell properly deprecates judicial ignorance of science, and is not hopeful that the 1993 US Supreme Court's opinion in the Daubert case, which established criteria for admissibility of scientific evidence, will be effective. The United States has 50 constitutionally sovereign state jurisdictions, each of which establishes its own evidentiary criteria, and which together handle a hundred times as many cases as the federal courts. Although most federal courts are at least beginning to require valid scientific evidence of causation, it will be years before all state courts adopt such standards, or other reforms.

Furthermore, class actions and unlimited punitive damages still provide ample opportunity for abuse; and a bill to limit such abuses was vetoed by President Bill Clinton last May.

Much has been written about implant cases, but mainly by journalists and lawyers. Angell presents a detailed, comprehensive, fairly balanced narrative from a scientific viewpoint. Yet the title is misleading — it is the law, rather than science, that is on trial in these cases. They teach that the alternative to science is not justice but ignorance; lacking scientific evidence of causation, litigation becomes a lottery.

Anyone concerned with the social impact of either science or law will find this book interesting and informative, as well as highly readable. □

Lee Loevinger is at Hogan and Hartson, 555 Thirteenth Street NW, Washington DC 20004-1109, USA.



GREAT apes can think at symbolic levels traditionally considered to be uniquely human. That is the claim of the field and laboratory researchers who contribute papers to *Reaching into Thought: The Minds of the Great Apes*, edited by Anne E. Russon, Kim A. Bard and Sue Taylor Parker Cambridge University Press, £55.

Goodness denied

Frans B. M. de Waal

The Origins of Virtue. By Matt Ridley. Viking: 1996. Pp. 295. £20. To be published in the United States in March 1997.

ANYONE who has watched the film *Il Postino* realizes the extraordinary lure of the metaphor. The apprentice poet learns to offer a fresh, new look at the world through carefully selected analogies. Shy at first, he soon relishes the poet's proverbial 'licence' to transform reality.

Scientists lack such a licence. Metaphors are used in science to great effect and advantage, but also at great peril. When metaphors are taken literally, they begin to obscure the truth. This lot befell the well-known struggle-for-existence view of the natural world. It kept generations of biologists from seeing the obvious conflux of interests among individuals and species even though Charles Darwin — always wiser than his followers — warned: "I use the term Struggle for Existence in a large and metaphorical sense including dependence of one being on another."

In *The Origins of Virtue*, Matt Ridley does an admirable job at documenting the mutual dependencies implied by Darwin, but not without getting seriously entangled in the field's latest metaphor according to which genes are selfish. Ridley, a British science writer who dazzled us, in 1993, with a well-received exposé of sexual evolution, *The Red Queen*, now reviews and elucidates the rich literature on tit-for-tat strategies, the prisoner's dilemma, group identity and other issues related to the evolution of altruism. The book is extremely well written, with the sort of anecdotal detail and wit that make for lively reading even

when the most abstract topics are being treated. It should be of interest to anyone following the debates about the origins of cooperation and morality.

Although the influence of George Williams and William Hamilton is compared to that of Copernicus and Darwin — surely a slight exaggeration — the book reads like a tribute to a third contemporary evolutionist: Robert Trivers. Trivers' ideas about reciprocal altruism are finally taking the central position in evolutionary accounts that they deserve; they are at the core of Ridley's argument about how 'virtue' might have come into existence. People enter into give-and-take deals, are concerned about their reputations within society and punish those who try to give less than they receive. We do unto others as we would have them do unto us; a rule that seems not too far removed from the way in which chimpanzees and some other animals selectively share food with those who share with them.

The author touches on a wide range of issues, from the growing interest of economists in commitments and fairness to the 'tragedy of the commons' — the over-exploitation resulting when people share resources without restrictions — and the social policy implications of evolutionary theory. Leaning to the conservative side, as represented by Newt Gingrich and Margaret Thatcher, he argues that we are doomed if we keep denying our species' inherent opportunism and egoism. Society would work a whole lot better with less government and more room for private enterprise driven by human greed. Like Thatcher (according to whom "there is no such thing as society"), Ridley is not much of a believer in the greater good.

Although the chapter containing these ideas is prefaced by a self-mocking comment about how "the author suddenly and rashly draws political lessons", readers

hardly need to be surprised. First, there is the systematic playing down of animal altruism: "[t]he essential virtuousness of human beings is proved not by parallels in the animal kingdom, but by the very lack of convincing animal parallels". We are urged not to bowdlerize nature by playing up the slimmest of clues to animal virtue — as if, ever since Konrad Lorenz and Robert Ardrey, the literature on animal behaviour has done exactly the opposite: that is, overemphasized the competitive and aggressive side. The postulated discontinuity between humans and animals serves to position morality as a recent advance. In this, the author follows a tradition going back to Thomas Henry Huxley's "Evolution and Ethics" lecture in 1893, which granted morality only the shallowest of roots in human nature.

Second, the book stresses that selfishness is in our genes, and that genes are our masters. Time and again, we receive seemingly contradictory explanations about the selfishness of genes versus the occasional unselfishness of the behaviour that these genes produce. Of course, every biologist believes that behaviour evolved because of the benefits associated with it, but it is essential to realize that for terms such as 'selfishness' or 'unselfishness' to have any meaning at all, behavioural effects ought to be recognizable to the actors, not just to evolutionary biologists. Most animal behaviour even human, is poorly categorized along this dimension.

Ironically, a teleological position left behind long ago with regard to the evolution of species has now achieved popularity with regard to genes. What the selfish-gene metaphor has done is infuse the evolutionary process with direction and intent. Genes are said to be in charge, and to strive for their own reproduction. But really all that is going on is that genes — a mere batch of molecules — replicate at different rates dependent on the traits they produce. Rather than doing the selecting themselves, they are *being* selected. As soon as the idea that they have a goal in life is abandoned, the paradox that Ridley tries to resolve evaporates before our eyes.

This critical reflection is absent, however. *The Origins of Virtue* reads like a conflation between the nonexistent psychology of genes and the actual psychology of individuals. Virtue can be found in the constructs of human sociality, such as the marketplace or bonds of trust, but is absent from the hearts and souls of people. No doubt some readers will agree with this view, but the book can also be read as a testimony to the outlived usefulness of yet another metaphor.

Frans B. M. de Waal is at the Yerkes Regional Primate Research Center and the Department of Psychology, Emory University, Atlanta, Georgia 30322, USA.

How the camel got his hump

Juliet Clutton-Brock

A Perfect Harmony. By Roger A. Caras. Simon and Schuster: 1996. Pp. 271. \$23.

THE behavioural and cultural changes brought about in animals by domestication have been a neglected area of study. This is in great part because domestic animals have been considered to be artefacts of human culture and therefore outside nature. But views are changing and the biological adaptations that have enabled certain species of animals to flourish in human societies can be seen to result from the same evolutionary process that produced the human species.

With the knowledge that there have been anthropogenic influences altering environments for upwards of 70,000 years, it is becoming meaningless to talk of wilderness areas. Soon, the necessity to manage wildlife will mean that the wild will merge with the tame. For this management to be successful, it is becoming imperative to study the interactions of humans and animals as they were in the past and to consider how they should be in the future.

Ethology, the study of animal behaviour, joined what *New Scientist* magazine recently called "the pantheon of 'true' sciences" in the 1960s. Archaeozoology, the study of animal remains from archaeological sites, providing evidence for the history of domestication, joined in the 1970s.

Anthrozoology, the study of interactions between humans and animals, is still struggling to push its way in, and recognition by the scientific community has been slow. This is despite the launch of a journal, *Anthrozoös* (reviewed in *Nature* **341**, 364; 1989), and the foundation of the International Society of Anthrozoology, which this year held a successful international conference in Cambridge, England.

Research in anthrozoology is burgeoning in the United States, with authoritative publications by the likes of Harriet Ritvo, Andrew Rowan and James Serpell. None of these authors is cited in *A Perfect Harmony*, which is not intended to be a report of recent research but is mostly an account of the author's personal views of the impact of animals on the history of civilization. Roger Caras is clearly most at ease when recounting anecdotes. These provide some intriguing bits of information, such as how the word 'trivia' — Latin for the junction of three roads — became associated with 'gossip'. This was because people met and gossiped at the markets traditionally held at such junctions.

Caras is not so happy with the dry

bones of fossil history and sometimes he goes hopelessly awry, for example where he writes: "Some Stone Age cultures without domestication did blossom, but they were quick and lonely affairs, like unseen firework displays. The world and all history were untouched by them." The book is, alas, full of misinformation of this kind, for, after all, the Palaeolithic period lasted around 100,000 years and in Europe it ended only some 12,000 years ago.

Caras claims that the ancestral species of domestic cattle has not been identified in fossil form in Europe. This is a strange comment, as the remains of *Bos primigenius* (known to be the ancestor of all taurine cattle) are commonly found on sites throughout the Palaearctic region.

Apart from these idiosyncratic interpretations, much of the known history of domestic animals is well summarized and interspersed with informative comments on every aspect of past and present interactions between humans and animals. The book is illustrated with charming drawings and has a foreword by the great futurist and friend of the author, Arthur C. Clarke.

It has to be said, however, that *A Perfect Harmony* is a peculiar title for a book about the exploitation of animals which have often suffered, and still suffer, great cruelty. Harmony is a concept that is surely quite alien to most relationships between humans and animals, but presumably it is the author's hope for the future, as in the Breughel painting of "Earthly Paradise" reproduced on the book's jacket.

The recognition that domestic animals are as worthy of study as wild species and discussions now being widely held on consciousness in animals owe much to the animal-welfare movement. It is therefore rather surprising that Caras, who is president of the American Society for the Prevention of Cruelty to Animals, does not have more to say about these topics. The inappropriate and often cruel management of farm animals is mentioned in an afterword, but surely there should have been a serious discussion of the animal as an individual and of the ethics of intensive farming.

Because it is very readable, *A Perfect Harmony* may join Stephen Budiansky's *The Covenant of the Wild* (Morrow, New York, 1992 — for a review see *Nature* **356**, 487; 1992) as a popular and much discussed book, but neither should be read as an authoritative text. □

Juliet Clutton-Brock, editor of the *Journal of Zoology*, lives at South Barn, High Street, Fen Ditton, Cambridge CB5 8ST, UK.

■ In the *Company of Animals: A Study of Human-Animal Relationships* by James Serpell is now out in paperback. Cambridge University Press, £7.95, \$11.95.

Functional diversity of helper T lymphocytes

Abul K. Abbas, Kenneth M. Murphy & Alan Sher

The existence of subsets of CD4⁺ helper T lymphocytes that differ in their cytokine secretion patterns and effector functions provides a framework for understanding the heterogeneity of normal and pathological immune responses. Defining the cellular and molecular mechanisms of helper-T-cell differentiation should lead to rational strategies for manipulating immune responses for prophylaxis and therapy.

THE concept of heterogeneity of immune responses evolved from early clinical observations on infections that presented with strikingly different and polarized clinical and pathological features. For instance, the healing (tuberculoid) form of leprosy was found to be associated with strong delayed-type hypersensitivity (DTH) reactions and low levels of antibody, whereas uncontrolled (lepromatous) leprosy was associated with high antibody titres and weak DTH¹. The inverse relationship between antibody and DTH could be experimentally manipulated by varying the dose and form of antigen used to immunize animals^{2,3}. These results suggested a surprising dichotomy in the nature of immune responses to natural infections and experimental immunization. A paradox emerged when it was established that both antibody and DTH reactions to protein antigens were dependent on the helper function of CD4⁺ T lymphocytes—how could the same helper cells be involved in such diverse immune responses?

The explanation for the diversity of CD4⁺ T-cell-dependent immune responses came with the application of two important techniques for analysing lymphocytes, namely methods for propagating cloned lines of T cells and assays for cytokines. In a seminal study, Mosmann *et al.* demonstrated that mouse CD4⁺ T-cell clones could be classified into distinct populations on the basis of their patterns of cytokine production⁴. They termed the two most readily discernible populations Th1 and Th2, with Th1 clones being the ones that produced interleukin(IL)-2, interferon(IFN)- γ and tumour-necrosis factor(TNF)- β , and Th2 clones those producing IL-4, IL-5, IL-6 and IL-13 (Fig. 1). Although IL-10 was originally described as a product of Th2 clones⁵, it is now clear that IL-10 is also secreted by Th1 cells (especially in humans)⁶ and by activated macrophages. Evidence rapidly accumulated showing that Th1 and Th2 clones differed in the types of antibody response they stimulated, that Th1 clones mediated DTH reactions, and that Th2 cells were more potent helpers for antibody production, thus providing a simple explanation for the observed dichotomy of immune responses^{7,8}. Two other features of these clones proved to be seminal in our understanding of the generation of helper-T-cell diversity. First, it was shown that each T-cell subset produces cytokines that serve as its own autocrine growth factor⁹ and, as we shall discuss later, promote differentiation of naive T cells to that subset. Second, the two subsets produce cytokines that cross-regulate each other's development and activity (Fig. 1). For instance, IFN- γ produced by Th1 cells amplifies Th1 development and inhibits proliferation of Th2 cells¹⁰, whereas IL-10 produced by Th2 cells blocks activation of Th1 cells⁵. The net result of cytokine-mediated self-amplification and cross-regulation is that once a T-cell immune response begins to develop along one pathway, namely Th1 or Th2, it tends to become progressively polarized in that direction.

In the ten years since the first description of helper-T-cell subsets, studying the frequencies of these populations and their roles in various diseases has become a major focus. It is now possible to propose testable models for the differentiation and

functions of these subsets and to exploit these concepts for developing prophylactic vaccines and therapeutic approaches for immunological diseases.

Heterogeneity of T-cell populations

Following the discovery of mouse Th1 and Th2 clones, numerous attempts were made to define the patterns of cytokine production in T cells stimulated in different ways, and in cells isolated from experimental animals and from humans during the course of normal and pathological immune responses. With the advent of sensitive assays for cytokine production in individual cells such as *in situ* hybridization and immunostaining, it has become apparent that many T cells cannot easily be classified into Th1 and Th2 subsets based on the original criteria. Individual cells may exhibit complex and quite heterogeneous patterns of cytokine production, with various combinations of IL-2, IL-4, IL-5 and IFN- γ , and often produce cytokines, such as IL-10 and TGF- β , that serve important functions but are not characteristic of either subset. T cells that produce various mixtures of cytokines have been called 'Th0'. This heterogeneity of cytokine patterns may be even greater in human T lymphocytes than in murine cells. In fact, early difficulties in identifying Th1 and Th2 cells among human T lymphocytes generated considerable scepticism about the generality of this concept of heterogeneity. Such results have called into question even the existence of T-cell subpopulations whose functional differences are attributable to their cytokine profiles¹¹.

Two important sets of findings indicate that subsets of Th1 and Th2 cells do exist, and the degree of polarization and heterogeneity of T lymphocytes may reflect the nature of the antigenic and environmental stimuli to which the cells have been exposed. First, many experimentally induced and naturally occurring immune responses show patterns of cytokine production and effector reactions that are clearly indicative of Th1 or Th2 dominance. This is particularly true of responses to persistent infections with microbes such as *Leishmania*, *Listeria*, mycobacteria and helminths^{12,13}, or responses to non-infectious persistent antigens, as in allergies and autoimmune diseases. Studies of murine *Leishmania major* infection were especially informative, because they showed, for the first time, that resistance and susceptibility to disease are attributable to specific Th1 and Th2 responses¹⁴. Leishmaniasis remains a paradigm for the functional dichotomy of helper T cells and the *in vivo* significance of this phenomenon (discussed in more detail later). Another important development was the demonstration that Th1 and Th2 populations can be identified among human T cells stimulated in culture or isolated from patients with a variety of chronic infectious or immunological diseases¹⁵. Impressively, Th1/Th2 cytokine patterns are seen in some immune reactions even when total tissue RNA is analysed, without any of the biases introduced by cell culture and cloning^{16,17}. Furthermore, patterns of macrophage activation and the production of different antibody isotypes during various immune responses also support the existence of

Th1 or Th2 cells *in vivo*. Second, as implied above, the degree of Th1/Th2 polarization increases with the chronicity of immune responses. Thus, mixed (Th0) cytokine patterns are most noticeable early after lymphocyte activation¹¹, and the clearest demonstrations of Th1 and Th2 subsets are in chronic disease states, in which the antigens are persistent and cannot easily be eliminated¹⁵. The same phenomenon is observed *in vitro*, where repeated antigenic stimulation of T cells expressing single transgenic antigen receptors leads to increasingly polarized and irreversibly committed Th1 and Th2 populations^{18,19}. Not surprisingly, therefore, the frequencies of polarized Th1 and Th2 cells, and their ease of detection, may depend to a great extent on the type of immune response being analysed.

It is not yet clear how a dominant cytokine pattern at the population level could co-exist with heterogeneity of individual cells, or even how much discrepancy exists between the polarized patterns of cytokine production that are detected in bulk populations of lymphocytes and the variations among cells²⁰. Surprisingly, even within cloned lines of Th0 cells, cytokine production is heterogeneous in individual cells²¹, suggesting that the transcription and expression of different cytokine genes are regulated independently. Such results emphasize the importance of defining the intracellular pathways that regulate the production of cytokines. The lack of definitive phenotypic markers (other than cytokine profiles) for Th1, Th2 and other subsets compounds the problem of accurately quantifying the heterogeneity of cell populations. Because of the observed variations in patterns of cytokine production in individual cells, some immunologists suggest that 'Th1-like' and 'Th2-like', or 'Th1-dominant' and 'Th2-dominant', are more accurate designations for the subsets.

T lymphocytes other than CD4⁺ helper cells may consist of subsets that are also distinguishable by their patterns of cytokine production. For instance, CD8⁺ cytolytic T lymphocytes may be induced to differentiate into IL-4- and IFN γ -producing subpopulations *in vitro*^{22,23}. The functional differences between these subsets are not defined yet. Both are known to mediate specific cytotoxicity, but it is possible that they differ in their ability to sustain protective immunity against viruses and other intracellular microbes. Similar subpopulations are known to exist among T cells expressing the $\gamma\delta$ antigen receptor²⁴.

Th1 and Th2 effector functions

Soon after the identification of helper-T-cell subsets, it was proposed that Th1 cells were primary responsible for cell-mediated immunity or DTH, and Th2 cells for humoral immunity^{7,8}. Nevertheless, it was clear from the early studies that IFN- γ , the signature Th1 cytokine, induces the switching of B lymphocytes to several IgG isotypes, and was thus obligatory for some humoral immune responses^{25,26}. In fact, the rather oversimplified classification of T-cell effector functions into cell-mediated and humoral immunity has been a source of considerable confusion and controversy.

A logical functional classification of T-cell subsets must be based on the known biological activities of the cytokines these T cells produce (Fig. 2). The principal Th1 effector cytokine, IFN- γ , has two key functions. First, it activates macrophages, enhancing their microbicidal actions. Second, IFN- γ stimulates the production of IgG antibodies which bind to high-affinity Fc γ receptors and complement proteins and are therefore the principal antibodies involved in the opsonization and phagocytosis of particulate microbes. IFN- γ -dependent isotypes are best defined in mice, and are mainly IgG2a and IgG3 (ref. 27). Their human homologues are probably IgG1 and IgG3, but so far there is no good evidence that the production of complement-fixing and opsonizing IgG antibodies in humans is Th1- or IFN γ -dependent. Nevertheless, it is likely that the principal function of Th1 cells is to elicit phagocyte-mediated defence against infections, because Th1 cytokines promote the ability of macrophages to both phagocytose and destroy microbes. Th1-dominant immune responses are often associated with inflammation and tissue injury, because

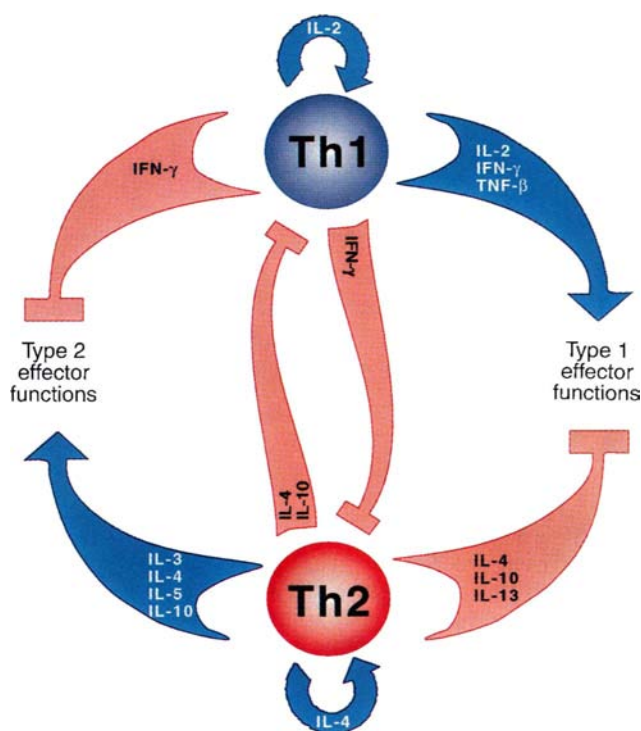


FIG. 1 The cytokines produced by Th1 and Th2 lymphocyte populations determine their opposing effector (blue) and inhibitory (pink) functions. The Th1 and Th2 pathways are symmetrical, each controlling a unique set of immune responses, and each augmenting the development of cells of the same subset while suppressing the expansion and/or effector functions of the other subset. IL-2 and IL-4 are shown as autocrine growth factors for Th1 and Th2 cells, respectively. The effector functions of Th1 and Th2 cells (referred to as type 1 and type 2 effector functions) are shown in more detail in Fig. 2.

two Th1 cytokines, TNF- β and IFN- γ , recruit and activate inflammatory leukocytes. The typical inflammatory reaction is DTH, and is often a price that is paid for protective immunity against microbes such as mycobacteria. Some Th1 cells acquire cytolytic capacity, and the cytokines produced by Th1 cells, notably IL-2 and IFN- γ , promote the differentiation of CD8⁺ T lymphocytes into active cytotoxic cells. These are the other ways in which the Th1 subset participates in the elimination of intracellular microbes.

The signature cytokines of Th2 cells are IL-4 and IL-5. IL-4 is the major inducer of B-cell switching to IgE production and is therefore a key initiator of IgE-dependent, mast-cell-mediated reactions²⁸. IL-5 is the principal eosinophil-activating cytokine²⁹, and mice lacking IL-5 or its receptor show marked defects in eosinophil responses to helminths^{30,31}. The production of these two cytokines by the same subset accounts for the frequent presence of both IgE and activated eosinophils in Th2-dominated immune reactions, such as in allergies and helminthic infections. The recruitment of eosinophils to sites of immune reactions is mediated by chemokines, including eotaxin, which are produced by both Th1 and Th2 subsets, and by non-T cells, and the activity of both is augmented by local IL-4 and IL-5 production^{32,33}. IgE and eosinophil-mediated defence is important for eliminating some but not all helminths³⁴, and it is still not clear how important Th2-induced, phagocyte-independent responses are in defence against other infectious agents. Th2 cells are excellent helpers for B lymphocytes, and stimulate the production of high levels of IgM and non-complement-fixing IgG isotypes, such as IgG1 in mice, or its homologue, IgG4, in humans²⁷.

It is noteworthy that several cytokines produced by Th2 cells have anti-inflammatory actions. IL-4 and IL-13 antagonize the macrophage-activating action of IFN- γ , IL-10 suppresses numerous macrophage responses, and TGF- β (which is produced by some Th2 cells and many other cell types) is antiproliferative and inhibits leukocyte activation. Thus, the net result of Th2 activation is to inhibit acute and chronic inflammation, including DTH reactions. This raises the possibility that an important physiological function of Th2 cells is not as effectors but as regulators of immune responses (Fig. 3). In particular, Th2 cells may appear late in immune responses, and serve to limit the injurious consequences of Th1-mediated protective immunity. Consistent with this hypothesis is the finding that during T-cell activation *in vitro*, Th1 responses often develop early and Th2 cells accumulate as the response progresses¹⁸. A similar regulatory function has been proposed for some CD8⁺ T cells that secrete 'type 2' cytokines³⁵.

Development of helper T-cell subsets

Studies using T cells stimulated with polyclonal activators and T cells from mice expressing transgenic antigen receptors of known specificities have established the key concept that Th1 and Th2 cells are not derived from distinct lineages that are committed before stimulation³⁶. Rather, both subsets may develop from the same T-cell precursor, whose differentiation is influenced by the manner and environment in which the precursors are stimulated (Fig. 4). This precursor is a mature, naive CD4⁺ T lymphocyte that produces mainly IL-2 upon initial encounter with antigen. The most potent differentiation-inducing stimuli are cytokines themselves. IL-12, produced by activated macrophages and dendritic cells, is the principal Th1-inducing cytokine. Potent adjuvants, which activate macrophages, and many microbial products,

including endotoxin and undefined components of viruses, intracellular bacteria such as *Listeria* and mycobacteria, and protozoa such as *Toxoplasma*, stimulate IL-12 production and induce Th1-dominated immune responses³⁷⁻³⁹. Among CD4⁺ lymphocytes, functional receptors for IL-12 appear to be restricted to recently activated, uncommitted cells and to Th1 cells, and are lost on differentiated Th2 cells⁴⁰. IL-12 activates three putative transcription factors, Stat1, Stat3 and Stat4 (ref. 41). Of these, Stat4 appears to be activated by IL-12 and no other cytokine, and knocking out either IL-12 (ref. 42) or Stat4 (refs 43, 44) results in markedly reduced Th1 responses. IFN- γ also promotes Th1 development, partly by enhancing IL-12 secretion by macrophages³⁹ and partly by maintaining the expression of functional IL-12 receptors on CD4⁺ T cells, rendering the cells more responsive to IL-12 (ref. 45). IFN- γ may have a minor Th1-inducing effect which is independent of IL-12. Thus, macrophages respond to many microbes by producing IL-12, which stimulates the development of the T-cell subset whose function is to enhance the microbicidal activities of the macrophages. There is obviously an elegant symmetry in this sequence of events.

The development of Th2 cells from naive precursors is induced by IL-4 (ref. 36), which signals through activation of Stat6 and a protein first identified as an insulin-response substrate, called IRS-2 (ref. 46, 47). The mechanism by which Stat6 induces IL-4 production and Th2 development is unclear. A consensus Stat6-binding site is present in the IL-4 promoter, and a multimer of this site is capable of stimulating the transcription of a reporter gene in response to IL-4 (ref. 48). These results suggest that Stat6 may activate IL-4 transcription in response to IL-4 itself, and thus play a role in IL-4-induced Th2 development. Consistent with this hypothesis are the findings that knockout of the IL-4 gene

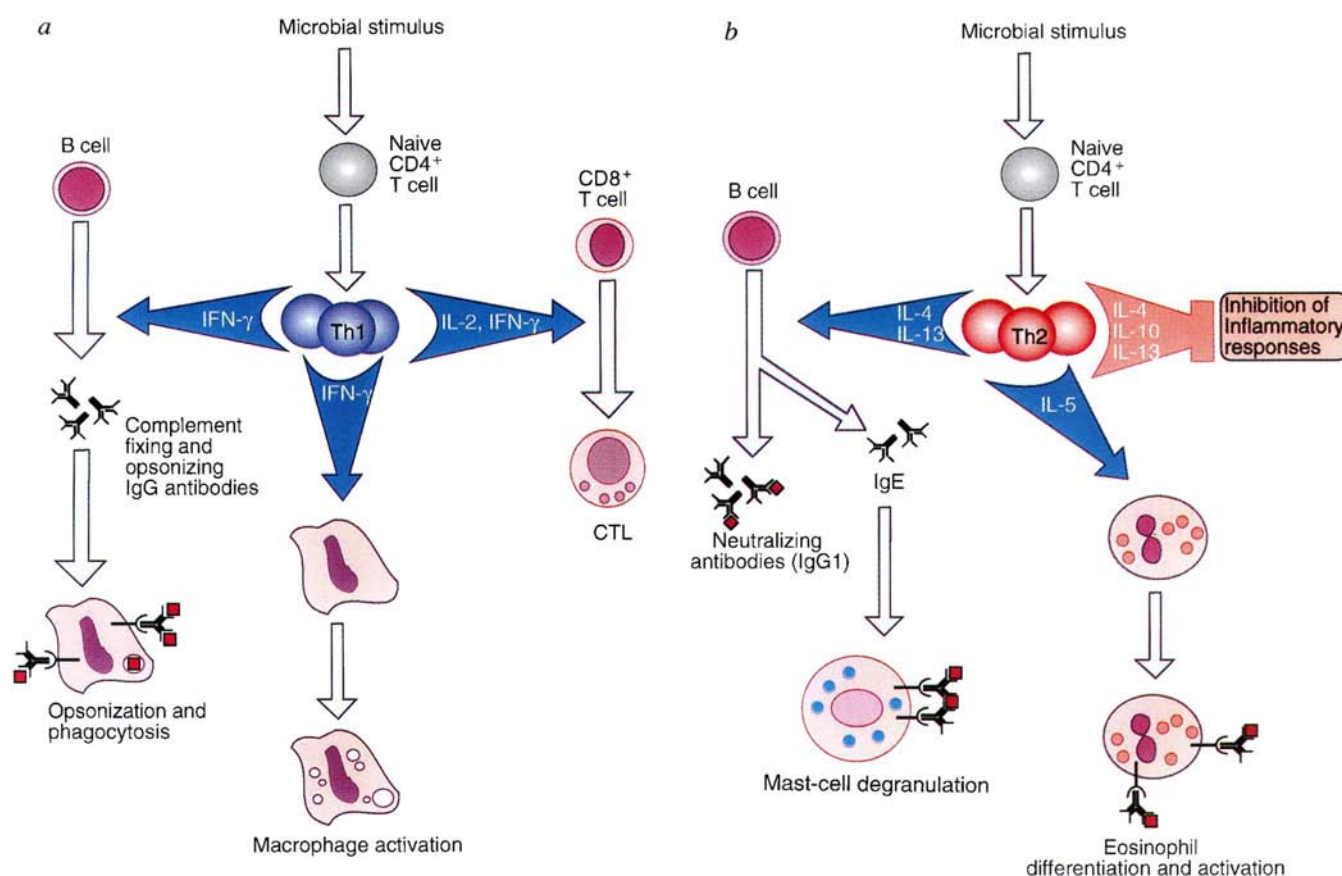


FIG. 2 Effector function of Th1 and Th2 subsets of CD4⁺ helper T lymphocytes. Th1 cells (a) induce phagocyte and T-cell-mediated defence reactions against microbes; Th2 cells (b) induce IgE-dependent mast-cell

degranulation and eosinophil activation, two components of the immediate hypersensitivity response.

(refs 49, 50) or the Stat6 gene⁵¹⁻⁵³ results in deficient Th2 responses, the converse of the IL-12 and Stat4 knockouts. The realization that IL-4 is required for Th2 development raised a conceptual problem, because it was thought that IL-4 would be produced by T cells only after their differentiation into Th2 effectors. It now appears likely that T cells themselves produce small amounts of IL-4 from their initial activation⁵⁴, and the amount of IL-4 that accumulates at the site of a T-cell response increases with increasing lymphocyte activation. The Th2-inducing effect of IL-4 dominates over other cytokines⁵⁷, so that if IL-4 levels reach a necessary threshold, Th2 differentiation is initiated and IL-4 production increases progressively. This explains why Th2 responses become increasingly pronounced with repeated T-cell stimulation, and why the magnitude of T-cell activation is an important determinant of Th2 development (see below). Under certain circumstances, other cells may produce IL-4 and direct Th2 development. One such source of early IL-4, identified in mice, is a small population of CD4⁺NK1.1⁺ cells which may recognize antigens presented in association with the non-polymorphic β_2 -microglobulin-associated molecule CD1 (ref. 55). However, the importance of these cells in *in vivo* Th2 responses to helminths and allergens is not yet established.

At least two other factors are important in determining the balance between Th1 and Th2 subsets in immune responses. The first is the dose or concentration of the antigen. With few exceptions, low antigen concentrations, and low-dose infections, tend preferentially to induce Th1 responses, whereas high doses induce Th2 development^{56,57}. The mechanisms underlying these effects of antigen dose are not well understood. It is possible that at low doses of antigen, the principal antigen-presenting cells (APCs) are dendritic cells, or macrophages if the antigen is a particulate microbe. Both dendritic cells and macrophages produce IL-12, and tilt the balance of the specific T-cell response towards Th1 differentiation. Conversely, when the antigen concentration is high, it may be presented by APCs that do not secrete IL-12, thus favouring Th2 development. It is also possible that high concentrations of antigen lead to repeated T cell stimulation, thus increasing IL-4 production and Th2 development, or induce a state of immunological tolerance (unresponsiveness), which often preferentially shuts off Th1 cells^{58,59}. Different concentrations of antigens may also trigger distinct signals from T-cell antigen receptors, which may influence the pattern of differentiation. Although this is an interesting hypothesis, so far there is no formal evidence to support it.

Another factor known to influence T-cell development is co-stimulation, which refers to signals provided by APCs that work together with antigen to enhance specific T-cell responses. The best-defined co-stimulators are two structurally related proteins, B7-1 and B7-2, both of which activate T cells by interacting with their specific receptor, CD28 (ref. 60). The development of both Th1 and Th2 cells is dependent on co-stimulation. In fact, high levels of co-stimulation promote Th2 responses, probably because increasing the magnitude of initial T-cell activation increases IL-4 production and the resultant IL-4-dependent autocrine pathway of Th2 differentiation. In contrast, Th1 development depends on IL-12 produced by APCs, probably working together with some level of co-stimulators⁶¹, and is less influenced by the magnitude of the T-cell response. It has also been shown in some systems that B7-1 and B7-2 differentially regulate the development of Th1 and Th2 cells^{62,63}. However, there is no clear evidence that these two co-stimulators deliver different biochemical signals to T cells. It appears more likely that the co-stimulator that influences a particular response depends on which one is being expressed, with B7-2 being dominant on unactivated APCs and B7-1 increasing after exposure of APCs to inflammatory stimuli⁶⁰. The issue of co-stimulation and effector T-cell development is further complicated by the possibility that different T-cell subsets vary in their dependence on co-stimulation at various stages of activation and differentiation. We have already emphasized the hypothesis that Th2 development may be enhanced by increasing the initial co-

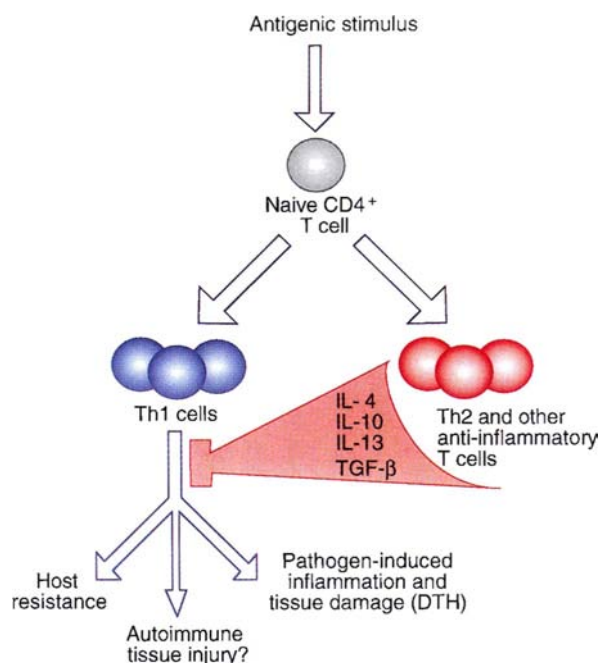


FIG. 3 The regulatory function of Th2 cells. Cytokines produced by Th2 cells control the inflammation and tissue injury that may accompany Th1-mediated immune reactions to infections, and may protect against autoimmune tissue damage. Excessive Th2-mediated suppression is associated with defective cellular immunity against intracellular microbes. DTH, delayed-type hypersensitivity.

stimulation by APCs. In contrast, fully differentiated Th2 cells respond to antigen without co-stimulation by B7, whereas Th1 cells continue to require B7 for their activation⁶⁴. Such results highlight the emerging view that the timing and level of co-stimulation, rather than simply its presence or absence, are critical determinants of the nature and magnitude of T-cell responses. Other surface molecules on T cells and APCs may also regulate the differentiation of the T cells. The interaction of CD40 ligand on T cells with CD40 on APCs is required for the development of effector T cells⁶⁵, but there is no evidence that this ligand: receptor pair influences the relative proportions of Th1 and Th2 cells that develop. Human Th2 cells express high levels of CD30 (ref. 66), but the role of this protein in the development of Th2 cells is unknown. There is also some evidence that antigen presentation by macrophages and dendritic cells preferentially induces Th1 responses, presumably because of concomitant IL-12 production, whereas antigen presentation by B cells stimulates Th2 development^{10,67}.

Taken together, these studies suggest that immunological stimuli that activate macrophages or dendritic cells and trigger IL-12 production tend to initiate Th1 responses. This would be characteristic of the adaptive immune response to many pathogenic microbes, especially intracellular bacteria and viruses. In contrast, antigens that are persistent or present at high concentrations and do not elicit large amounts of IL-12 would tend to induce Th2 responses. These features are typical of helminthic parasites and probably of environmental allergens. Once antigen-stimulated T cells begin to differentiate along a particular pathway, the cytokines they produce amplify their growth and development and suppress the reciprocal pathway (Fig. 1). Therefore, it is likely that the ultimate dominance of Th1 or Th2 cells in any immune response is due to a combination of initial IL-12- or IL-4-induced differentiation, which is associated with selective transcription of cytokine genes, and amplification or suppression of either subset of differentiating cells.

Genetic regulation of Th1/Th2 development. The most striking example of genetic control of Th1/Th2 development is the resistance or susceptibility of inbred mice to infection with *Leishmania major*, which correlates with specific Th1 or Th2 responses, respectively¹⁴. The control of resistance to *L. major* is complex and probably multigenic. *In vitro* studies indicate that CD4⁺ T cells from two mouse strains that are resistant and susceptible to *L. major*, B10.D2 and BALB/c, respectively, differ in their intrinsic tendency to develop towards Th1 or Th2 cells even in response to model protein antigens (other than *L. major*). Under neutral conditions of *in vitro* differentiation, namely antigen stimulation without added IL-12 or IL-4, BALB/c T cells acquire a more Th2-like phenotype and lose IL-12 responsiveness more rapidly than do B10.D2 cells⁶⁸. This difference is attributable to greater initial IL-4 production by BALB/c T cells and more prolonged maintenance of IL-12 responsiveness in B10.D2 cells⁴⁵. The molecular basis of these effects is not yet defined, but initial results suggest they may be controlled by one dominant genetic locus (K.M.M., unpublished data).

Stability of differentiated Th1 and Th2 populations. A question of considerable biological and practical importance is whether differentiated Th1 and Th2 populations are irreversibly committed, or whether interconversions are possible. Several experimental systems have shown that early in their development, both Th1 and Th2 populations may be interconvertible, but with repeated stimulation both populations become irreversibly committed^{6,19,69}. The biochemical basis of this reversibility or stability may be the regulated transcription of cytokine genes or the expression of cytokine receptors. For instance, some studies indicate that Th1 populations generated by one round of *in vitro* stimulation can readily be converted into IL-4 producers by culture with IL-4, but that IL-4-producing Th2 cells are stable even upon culture with IL-12 (ref. 69). The reason for this is that early Th1 cells retain their responsiveness to IL-4, but early Th2 cells lose expression of the signalling chain of the IL-12 receptor and therefore their ability to activate Stat4 in response to IL-12 (ref. 40). There may be conditions in which even Th2 populations retain functional IL-12 receptors, for example if they are exposed to IFN- γ , which maintains expression of the IL-12 receptor β -chain (K.M.M., manuscript in preparation). Under these conditions, developing Th2 cells may be capable of subsequent conversion into IFN- γ producers. Thus, the stability or reversibility of differentiated effector populations of helper T cells may be influenced to a large extent by the conditions of initial priming. A major limitation of studies examining the stability of Th1 and Th2 cells is that these have been done with cell populations and not with single cells. It is therefore unclear whether any observed phenotypic reversions are due to true conversions of differentiated cells or to development of effectors from uncommitted cells in the population. The ability selectively to manipulate differentiated effector T cells will be important for controlling pathological T-cell responses.

Despite extensive investigation into the development of Th1 and Th2 populations, the biochemistry of this process remains poorly understood. Other than cytokine-induced STAT pathways, little is known about the signals that drive commitment to either subset. It is possible that the pattern of helper-T-cell commitment may be influenced by the nature of the antigen-induced signal, which in turn may depend on the affinity of antigen binding to T-cell receptors, or the density of antigen on APCs. Such qualitative variations in antigen-receptor-generated stimuli could regulate Th1/Th2 development if Th1 and Th2 cytokine genes were responsive to distinct stimuli. At present we have no formal proof to support this idea. There is some evidence that the antigen receptors on Th1 and Th2 cells transduce different signals as a consequence of antigen binding. For example, Th1 and Th2 clones differ in the magnitude of Ca²⁺ mobilization following antigen-receptor crosslinking¹⁰. Furthermore, tolerogenic proteins^{58,59} and 'altered peptide ligands'⁷⁰, which are antigens bearing mutations in their T-cell-receptor contact residues, tend to stimulate Th2 development and preferentially block Th1 responses. These are

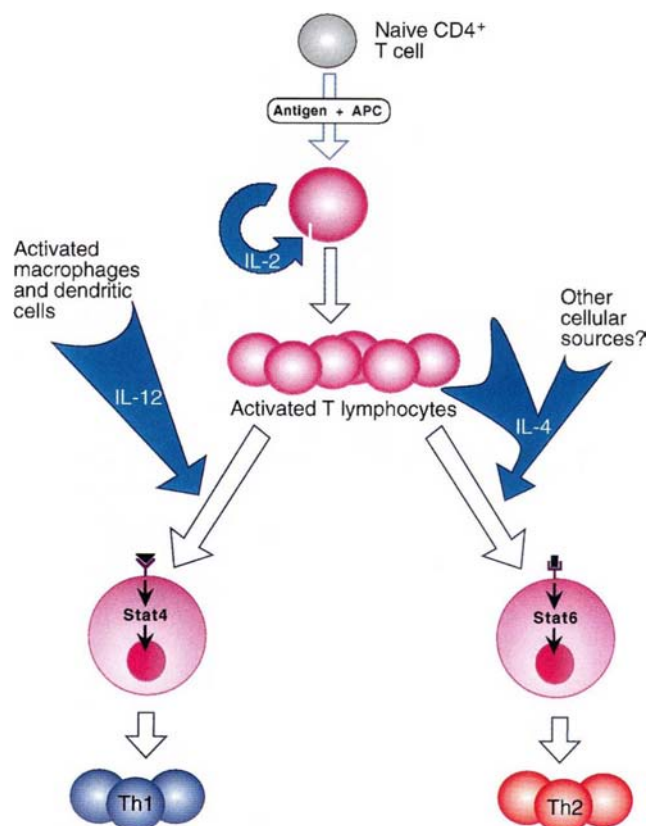


FIG. 4 The roles of cytokines in the development of Th1 and Th2 subsets. IL-12 and IL-4 activate distinct STAT proteins in antigen-stimulated T cells, and induce the differentiation of these cells to Th1 and Th2 populations. The principal sources of IL-12 are activated macrophages and dendritic cells, whereas IL-4 is produced by antigen-stimulated T cells themselves; other IL-4-producing cell types (for example, CD4⁺NK1.1⁺ T cells, mast cells and basophils) may also contribute to Th2 induction.

intriguing results, but neither their biochemical basis nor their physiological significance is understood.

T-cell subsets in disease

An important aspect of the Th1/Th2 dichotomy is the insight this concept provides into immunological determinants of disease. Indeed, the outcomes of a wide range of pathological processes, including infectious, allergic and autoimmune disorders, have been linked to Th1- or Th2-like cytokine expression patterns, and, by inference, to the particular T-cell subset induced. Such associations underscore the dual nature of T-cell-dependent effector and regulatory mechanisms, as well as the critical roles played by cytokines in controlling pathological immune responses.

Infectious diseases. The balance between Th1 and Th2 responses to an infectious agent can influence both pathogen growth and immunopathology. As mentioned earlier, studies of cutaneous leishmaniasis in inbred mice provided the first clear demonstration that resistance and susceptibility to an infectious disease correlated with antimicrobial Th1 and Th2 responses, respectively¹⁴. It is now appreciated that resistance to many intracellular microbes, including bacteria, protozoa and fungi, is linked to the induction of Th1 responses, and, in particular, with the macrophage-activating cytokines, IFN- γ and TNF- α ^{12,13}. Whereas Th1 responses also predominate in most viral infections, a more diverse set of effector mechanisms, including IFN- α/β -activated natural killer (NK) cells, cytolytic CD8⁺ T lymphocytes and neutralizing antibodies (often with Th1 isotype patterns), are

responsible for host resistance⁷¹. In many of these intracellular infections, IL-12 is induced early in the infection, and plays a central role in establishing IFN- γ -dependent resistance, as well as polarized Th1 response profiles^{39,72}. The mechanism by which macrophages and other leukocytes respond to intracellular pathogens by producing IL-12 is currently under active investigation.

Antimicrobial Th1 responses can also result in pathology, in the form of granulomatous inflammation, arthritis and colitis¹⁵. In many instances, the host tissue damage stems from toxic side effects of cytokines and other inflammatory mediators released during the normal immune attack on the pathogen. Severe pathological reactions may also result from defective cross-regulation by IL-10, TGF- β and other cytokines that normally inhibit Th1 effector functions (Fig. 3). This is dramatically illustrated by the increased susceptibility of IL-10-knockout mice to T-cell-dependent bacterial and protozoan-induced inflammatory disease^{73,74}. The resemblance of these pathologies to autoimmune disease (discussed later) supports the notion of a common mechanism of tissue injury due to excessive or uncontrolled Th1 responses.

In contrast to intracellular microbes, extracellular pathogens and particularly parasitic helminths typically trigger Th2-dominated responses¹². Th2-dependent IgG antibodies may contribute to the neutralization of toxins produced by extracellular bacteria. However, there is considerable controversy about the protective role of the prominent Th2-induced IgE, eosinophil and mast-cell responses that accompany metazoan worm infections. This is particularly so in mouse models, where ablation of eosinophil or IgE responses often has no effect on resistance to helminths. IL-4 is required for the control of several intestinal worm infections, apparently through induction of physiological changes in the gut, rather than through the action of IgE antibodies³⁴. It is also possible that helminths have evolved the capacity to induce Th2 responses to protect themselves against potentially toxic Th1-dependent antiparasite effector mechanisms¹¹. Regardless of the mechanisms and protective value of antihelminthic Th2 responses, it is clear that such responses may also be detrimental to the host, contributing to granuloma formation and hypereosinophilia^{15,75}.

A causal relationship between relative Th1/Th2 activity and the progression of infectious diseases in humans is suggested by findings in leprosy, in which tuberculoid and lepromatous lesions express predominant Th1 and Th2 cytokines, respectively¹⁶. A similar mechanism has been proposed to underlie the progression of AIDS⁷⁶. An early loss of IL-2, IFN- γ and IL-12 production and responsiveness is observed in HIV-1 infection^{76,77}, but there is considerable debate about whether there is a simultaneous increase in Th2 cytokine production and, if so, whether this explains both Th1 downregulation and viral progression^{15,78}. Perhaps more pertinent is the observation that HIV-1 preferentially infects Th2 clones, a finding that may explain the persistence of the virus in a Th1-deficient environment⁷⁹.

Allergic diseases. Allergic reactions involving IgE and mast cells are due to the development and activation of allergen-specific Th2 cells^{15,17}. Individuals with severe atopy have high levels of total and allergen-specific serum IgE. In some of these patients, there is a strong correlation between serum IgE and frequencies of allergen-specific Th2 cells that can be propagated from peripheral blood lymphocytes¹⁵. Many attempts are ongoing to define the genetic control of IgE-mediated allergy. One study of atopic patients suggests that the levels of serum IgE are controlled by a locus within chromosome band 5q31.1 (ref. 80). This region contains a cluster of cytokine genes, including the genes encoding IL-3, IL-4 and IL-5, as well as genes encoding signalling molecules that may be relevant to Th1/Th2 differentiation, such as the IFN- γ -induced protein IRF-1. Interestingly, preliminary analysis of the genetic control of the propensity of B10.D2 and BALB/c mouse T cells to develop into Th1 or Th2 subsets, respectively, suggests that the relevant locus maps to the syntenic region in mouse chromosome 11 (K.M.M., manuscript in preparation). Identification of

the gene(s) responsible will provide valuable insight into the mechanisms responsible for the development of allergic diseases and, even more broadly, into the control of T-lymphocyte differentiation pathways.

Self-tolerance and autoimmunity. Most studies of immunological tolerance have focused on the induction of tolerance to foreign protein antigens or to tissue transplants. Helper-T-cell tolerance to protein antigens can be induced by administering large doses of the antigens without adjuvants, by blocking co-stimulators during initial antigen recognition, or by administering mutated peptides that engage specific T-cell antigen receptors in an abnormal way⁸¹. In many of these experimental systems of peripheral tolerance, the induction and responses of Th1 cells are preferentially blocked. The finding that Th2 responses are often unaffected raises the possibility that the state of immunological unresponsiveness is at least partly due to the immunosuppressive cytokines produced by residual Th2 (or other regulatory) T cells^{82,83}. Evidence for this hypothesis has been obtained in some, but not all, experimental systems^{59,84}. Tolerance to self antigens may be mimicked by expressing various proteins such as allogeneic major histocompatibility complex (MHC) molecules or viral proteins in tissues, using transgenic technology⁸⁵. In these situations, the expressed protein is present throughout the development of the animal and is effectively a 'self' antigen. Here, too, tolerance is often associated with a block in the development of 'self'-antigen-reactive Th1 cells. Conversely, the activation of proinflammatory Th1 cells correlates with the induction of autoimmune tissue injury^{86,87}, and in some models shows the same genetic control as resistance to leishmaniasis⁸⁶.

A question that arises from these studies is whether self antigens normally elicit Th2 responses, and whether the anti-inflammatory cytokines produced by Th2 cells protect tissues from destructive autoimmune attack (Fig. 3). The protective cells may be T cells that secrete large amounts of TGF- β , rather than classical Th2 cells. Although this is an intriguing hypothesis, there is no convincing evidence that normal individuals make protective T-cell responses against their own tissue antigens. There is also no evidence that a decline of self-antigen-reactive regulatory T cells, and a resultant loss of control on anti-self Th1 cells, is a factor in any spontaneous autoimmune disease. It is worth mentioning that some antibody-mediated autoimmune diseases may be caused by Th2-induced antibodies, but these antibodies do not fix complement or bind to phagocyte Fc receptors and are unlikely to cause severe tissue injury.

Discussion

The evidence linking T-cell subsets to disease susceptibility and expression forms the basis of several immunological approaches for treatment and prophylaxis of such disorders. Indeed, there are now numerous examples of experimental models in which modulation of the Th1/Th2 balance by administration of recombinant cytokines or cytokine antagonists alters the outcome of disease. For instance, IL-12 administered at the time of infection enhances resistance to many intracellular protozoan, bacterial and fungal pathogens and to some viruses³⁹. When used as a vaccine adjuvant with sensitizing doses of antigen, IL-12 converts the recall response to challenge infection from a Th2 to a Th1 pattern, thereby promoting resistance to intracellular infection⁸⁸ while suppressing Th2-dependent pathology⁸⁹. In some situations, IL-12 even converts established Th2 responses to Th1 dominance, suggesting its possible application in the treatment of allergy^{89,90}. This cytokine also has potent effects against tumours in several experimental models⁹¹ and is being tested in anticancer vaccine protocols.

In the opposite direction, the inhibition of Th1 responses or the induction of Th2 cells is a potential approach for the treatment of inflammatory diseases. For example, the macrophage- and Th1-inhibitory actions of IL-10 have been exploited to suppress lipopolysaccharide-induced endotoxaemia⁹² and as a therapy for inflammatory bowel disease⁹³ in experimental animals. The

selective induction of Th2 cells has been proposed as a treatment for tissue autoimmune diseases⁸³.

The direct administration of cytokines and cytokine antagonists carries the risk of unwanted side effects. In addition, it is difficult to predict the detrimental effects of pharmacologically induced switching of T cells to either polarized subset. An alternative to cytokine therapy is to alter the endogenous Th1/Th2 balance by particular treatments. For example, oral administration of antigens favours the development of T cells that produce TGF- β and Th2 cytokines⁹⁴. This is the basis for 'oral tolerance' protocols for preventing autoimmune diseases.

The diversity of T-lymphocyte function revealed by studies of T-cell subsets is now a well documented feature of the mammalian immune system. Learning how optimally to exploit this functional dichotomy for practical applications will require greater knowledge of the mechanisms that regulate the development and effector functions of these subsets. We still do not fully understand how different doses or forms of antigens, and their routes of administration, preferentially induce responses dominated by one or the other T-cell subset. This is particularly true for the Th2 pathway, where both the early sources of IL-4 required for Th2

differentiation and the distinguishing biochemical features of the microbial antigens and allergens that trigger such responses are unknown. The intracellular biochemical signals that drive a single T cell towards either subset, and the basis of the genetic predisposition of individuals to mount responses dominated by one subset, are not defined. The role of the Th1/Th2 balance in the recognition of self antigens, and the importance of anti-inflammatory T cells for the maintenance of self tolerance, are also fundamental and unresolved issues. It is, however, remarkable that in the short time since the original description of T-cell subset dichotomy, this concept has become a major theme of immunology, and dominates research on immune regulation in health and disease. The technology is now in place to answer many of the unresolved questions and to exploit fully the concept of functional T-cell diversity for treating human diseases. □

Abul K. Abbas is in the Immunology Research Division, Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115; Kenneth M. Murphy is in the Department of Pathology, Washington University School of Medicine, St Louis, Missouri 63110; and Alan Sher is in the Immunobiology Section, Laboratory of Parasitic Diseases, NIAID, NIH, Bethesda, Maryland 20892, USA.

1. Turk, J. L. & Bryceon, A. D. M. *Adv. Immunol.* **13**, 209–270 (1971).
2. Asherson, G. L. & Stone, S. H. *Immunology* **9**, 205–217 (1965).
3. Parish, C. R. *Transplant. Rev.* **13**, 35–66 (1972).
4. Mosmann, T. R., Cherwinski, H., Bond, M. W., Giedlin, M. A. & Coffman, R. L. *J. Immunol.* **136**, 2348–2357 (1986).
5. Fiorentino, D. F., Bond, M. W. & Mosmann, T. R. *J. Exp. Med.* **170**, 2081–2095 (1989).
6. Sornasse, T., Larenas, P. V., Davis, K. A., de Vries, J. E. & Yssel, H. *J. Exp. Med.* **184**, 473–483 (1996).
7. Mosmann, T. R. & Coffman, R. L. *Annu. Rev. Immunol.* **7**, 145–173 (1989).
8. Bottomly, K. *Immunol. Today* **9**, 268–274 (1988).
9. Lichtman, A. H., Kurt-Jones, E. A. & Abbas, A. K. *Proc. Natl Acad. Sci. USA* **84**, 824–827 (1987).
10. Fitch, F. W., McKisick, M. D., Lancki, D. W. & Gajewski, T. F. *Annu. Rev. Immunol.* **11**, 29–48 (1993).
11. Kelso, A. *Immunol. Today* **12**, 374–379 (1995).
12. Sher, A. & Coffman, R. L. *Annu. Rev. Immunol.* **10**, 385–409 (1992).
13. Kaufman, S. H. E. *Annu. Rev. Immunol.* **11**, 129–163 (1993).
14. Reiner, S. L. & Locksley, R. M. *Annu. Rev. Immunol.* **13**, 151–177 (1995).
15. Romagnani, S. *Annu. Rev. Immunol.* **12**, 227–257 (1994).
16. Yamamura, M. et al. *Science* **254**, 277–280 (1991).
17. Robinson, D. S. et al. *N. Engl. J. Med.* **326**, 298–304 (1992).
18. Croft, M. & Swain, S. L. *J. Immunol.* **154**, 4269–4281 (1995).
19. Murphy, E. et al. *J. Exp. Med.* **183**, 901–913 (1996).
20. Openshaw, P. et al. *J. Exp. Med.* **182**, 1357–1367 (1995).
21. Bucy, R. P. et al. *J. Exp. Med.* **180**, 1251–1262 (1994).
22. Croft, M., Carter, L., Swain, S. L. & Dutton, R. W. *J. Exp. Med.* **180**, 1715–1728 (1994).
23. Sad, S., Marcotte, R. & Mosmann, T. R. *Immunology* **2**, 271–279 (1995).
24. Ferrick, D. A. et al. *Nature* **373**, 255–257 (1995).
25. Snapper, C. M. & Paul, W. E. *Science* **236**, 944–947 (1987).
26. Stevens, T. L. et al. *Nature* **334**, 255–258 (1988).
27. Coffman, R. L., Leberman, D. A. & Rothman, P. *Adv. Immunol.* **54**, 229–270 (1993).
28. Galli, S. J. *N. Engl. J. Med.* **328**, 257–265 (1993).
29. Wardlaw, A. J., Moqbel, R. & Kay, A. B. *Adv. Immunol.* **60**, 151–266 (1995).
30. Kopf, M. et al. *Immunity* **4**, 15–24 (1996).
31. Yoshida, T. et al. *Immunity* **4**, 483–494 (1996).
32. Gonzalo, J. A. et al. *Immunity* **4**, 1–14 (1996).
33. Rothenberg, M. E., Luster, A. D. & Leder, P. *Proc. Natl Acad. Sci. USA* **92**, 8960–8964 (1995).
34. Finkelman, F. D. et al. *Annu. Rev. Immunol.* (in the press).
35. Bloom, B. R., Modlin, R. L. & Salgame, P. *Annu. Rev. Immunol.* **10**, 453–488 (1992).
36. Seder, R. A. & Paul, W. E. *Annu. Rev. Immunol.* **12**, 635–673 (1994).
37. Hsieh, D. S. et al. *Science* **260**, 547–549 (1993).
38. Gazzinelli, R. T., Hieny, S., Wynn, T. A., Wolf, S. & Sher, A. *Proc. Natl Acad. Sci. USA* **90**, 6115–6119 (1993).
39. Trinchieri, G. *Annu. Rev. Immunol.* **13**, 251–276 (1995).
40. Szabo, S., Jacobson, N. G., Dighe, A. S., Gubler, U. & Murphy, K. M. *Immunity* **2**, 665–675 (1995).
41. Jacobson, N. G. et al. *J. Exp. Med.* **181**, 1755–1762 (1995).
42. Magram, J. et al. *Immunity* **4**, 471–481 (1996).
43. Thierfelder, W. E. et al. *Nature* **382**, 171–174 (1996).
44. Kaplan, M. H., Sun, Y.-L., Hoey, T. & Grusby, M. J. *Nature* **382**, 174–177 (1996).
45. Guler, M. et al. *Science* **271**, 984–987 (1996).
46. Hou, J. et al. *Science* **265**, 1701–1706 (1994).
47. Ryan, J. J. et al. *Immunity* **4**, 123–132 (1996).
48. Lederer, J. A. et al. *J. Exp. Med.* **184**, 397–406 (1996).
49. Kuhn, R., Rajewsky, K. & Müller, W. *Science* **254**, 707–710 (1991).
50. Kopf, M. et al. *Nature* **362**, 245–248 (1993).
51. Kaplan, M. H., Schindler, U., Smiley, S. T. & Grusby, M. J. *Immunity* **4**, 313–319 (1996).
52. Takeda, K. et al. *Nature* **380**, 627–630 (1996).
53. Shimoda, K. et al. *Nature* **380**, 630–633 (1996).
54. Schmitz, J. et al. *J. Exp. Med.* **179**, 1349–1353 (1994).
55. Yoshimoto, T. & Paul, W. E. *J. Exp. Med.* **179**, 1285–1295 (1994).
56. Bretscher, P. A., Wei, G., Menon, J. N. & Bielefeldt-Olmann, H. *Science* **257**, 539–542 (1992).
57. Hosken, N. A., Shibuya, K., Heath, A. W., Murphy, K. M. & O'Garra, A. *J. Exp. Med.* **182**, 1579–1584 (1995).
58. DeWit, D. et al. *J. Exp. Med.* **175**, 9–15 (1992).
59. Burstein, H. J. & Abbas, A. K. *J. Exp. Med.* **177**, 457–463 (1993).
60. Lenschow, D. J., Walunas, T. L. & Bluestone, J. A. *Annu. Rev. Immunol.* **14**, 233–258 (1996).
61. Murphy, E. E. et al. *J. Exp. Med.* **180**, 223–231 (1994).
62. Kuchroo, V. K. et al. *Cell* **80**, 707–718 (1995).
63. Lenschow, D. J. et al. *J. Exp. Med.* **181**, 1145–1155 (1995).
64. McKnight, A. J., Perez, V. L., Shea, C. M., Gray, G. S. & Abbas, A. K. *J. Immunol.* **153**, 5220–5225 (1994).
65. Grewal, I. S., Xu, J. & Flavell, R. A. *Nature* **378**, 617–620 (1995).
66. Romagnani, S. et al. *J. Leuk. Biol.* **57**, 726–730 (1995).
67. Mason, D. J. *J. Exp. Med.* **183**, 717–719 (1996).
68. Hsieh, C. S., Macatonia, S. E., O'Garra, A. & Murphy, K. M. *J. Exp. Med.* **181**, 713–721 (1995).
69. Perez, V. L., Lederer, J. A., Lichtman, A. H. & Abbas, A. K. *Int. Immunol.* **7**, 869–875 (1995).
70. Sloan-Lancaster, J. & Allen, P. M. *Annu. Rev. Immunol.* **14**, 1–27 (1996).
71. Biron, C. A. *Curr. Opin. Immunol.* **6**, 530–538 (1994).
72. Biron, C. A. & Gazzinelli, R. T. *Curr. Opin. Immunol.* **7**, 485–496 (1995).
73. Kuhn, R., Lohler, J., Rennick, D., Rajewsky, K. & Müller, W. *Cell* **75**, 263–270 (1993).
74. Gazzinelli, R. T. et al. *J. Immunol.* **157**, 798–805 (1996).
75. Wynn, T. A. & Cheever, A. W. *Curr. Opin. Immunol.* **7**, 505–511 (1995).
76. Clenci, M. & Shearer, G. M. *Immunol. Today* **14**, 107–111 (1993).
77. Chehimi, J. et al. *J. Exp. Med.* **179**, 1361–1366 (1994).
78. Graziosi, C. et al. *Science* **265**, 248–252 (1994).
79. Maggi, E. et al. *Science* **265**, 244–248 (1994).
80. Marsh, D. G. et al. *Science* **264**, 1152–1156 (1994).
81. Kruisbeek, A. M. & Amsen, D. *Curr. Opin. Immunol.* **8**, 233–244 (1996).
82. Liblau, R. S., Singer, S. M. & McDevitt, H. O. *Immunol. Today* **16**, 34–38 (1995).
83. Rocken, M., Racke, M. & Shevach, E. *Immunol. Today* **17**, 225–231 (1996).
84. Davies, J. D., Leong, L. Y., Mellor, A., Cobbold, S. P. & Waldmann, H. *J. Immunol.* **156**, 3602–3607 (1996).
85. Miller, J. F. A. P. & Flavell, R. A. *Curr. Opin. Immunol.* **6**, 892–899 (1994).
86. Scott, B. et al. *Immunity* **1**, 73–83 (1994).
87. von Herrath, M. G., Guerdier, S., Lewicki, H., Flavell, R. A. & Oldstone, M. B. *Immunity* **3**, 727–738 (1995).
88. Afonso, L. C. et al. *Science* **263**, 235–237 (1994).
89. Wynn, T. A., Eltoum, I., Oswald, I. P., Cheever, A. W. & Sher, A. *J. Exp. Med.* **179**, 1551–1561 (1995).
90. Gavett, S. H. et al. *J. Exp. Med.* **182**, 1527–1536 (1995).
91. Brunda, M. J. et al. *J. Exp. Med.* **178**, 1223–1230 (1993).
92. Howard, M., Muchamel, T., Andrade, S. & Menon, S. *J. Exp. Med.* **177**, 1205–1208 (1993).
93. Powrie, F. et al. *Immunity* **1**, 553–562 (1994).
94. Weiner, H. L. et al. *Annu. Rev. Immunol.* **12**, 809–837 (1994).

ACKNOWLEDGEMENTS. We thank M. Doherty for preparing the illustrations.

Origin of floral asymmetry in *Antirrhinum*

Da Luo, Rosemary Carpenter, Coral Vincent, Lucy Copsey & Enrico Coen

Genetics Department, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK

Dorsoventral asymmetry in flowers is thought to have evolved many times from a radially symmetrical ancestral condition. The first gene controlling floral asymmetry, *cycloidea* in *Antirrhinum*, has been isolated. The *cycloidea* gene is expressed at a very early stage in dorsal regions of floral meristems, where it affects growth rate and primordium initiation. Expression continues through to later stages in dorsal primordia to affect the asymmetry, size and cell types of petals and stamens.

FLOWERS can be classified into two basic types according to their symmetry: irregular flowers have a single plane of symmetry, whereas regular flowers have two or more planes of symmetry¹. Both flower types have a radial axis that establishes concentric whorls of different types of organs, typically sepals, petals, stamens and carpels^{2,3}. Irregular flowers have an additional axis of asymmetry such that organs from the same whorl have distinct identities according to their dorsoventral position⁴. The irregular condition is thought to have evolved many times from the regular one as a specialized adaptation to animal pollinators. One approach to understanding the origin and mechanisms underlying dorsoventral asymmetry is through the analysis of mutants in which asymmetry is reduced or lost. Such mutations have been known since the classic study on peloric forms of *Linaria* described by Linnaeus in 1744 (ref. 5), but so far they have not been the subject of detailed developmental or molecular analysis. We have chosen to address this problem in *Antirrhinum majus*, a species with irregular flowers that is amenable to molecular genetics.

The wild-type *Antirrhinum* flower has an axis of dorsoventral asymmetry which is orientated such that the dorsal (upper or adaxial) part is nearer the stem, whereas the ventral (lower or abaxial) part is nearer to the bract, a small leaf-like organ that subtends each flower. Dorsoventral asymmetry is most pronounced in the petals and stamens, each of which can be divided into three types: dorsal, lateral and ventral. Mutants with a peloric phenotype have radially symmetrical flowers in which all organs have a ventral identity, indicating that they lack genetic functions that are normally active in the dorsal and lateral regions of the flower^{4,6}. The semipeloric phenotype is intermediate between peloric and wild type. Two groups of semipeloric mutants were identified, mapping to two linked loci⁶. To avoid possible confusion between the two loci, here we call one locus *cycloidea* (*cyc*) and the other *radialis* (*rad*), names given to some of the original mutants identified in the classical genetic studies¹⁵.

We describe the isolation of *cyc* by transposon tagging and show that it is expressed in a dorsal domain of wild-type floral meristems, before any morphological dorsoventral asymmetry in the meristem is visible. The phenotypic consequences of its expression can be divided into early effects on primordium initiation and later effects on organ morphology. Early expression of *cyc* retards growth rate and reduces organ number in the dorsal region of the floral meristem, perhaps helping to ensure that the domain of *cyc* expression is aligned precisely with the arrangement of primordia in the flower. At later stages, *cyc* continues to be strongly expressed in dorsal organ primordia, where it interacts with the radial organ identity genes to affect the asymmetry, size and cell-types of petals and stamens. No expression of *cyc* is detected in lateral organs, suggesting that it may act non-autonomously to influence lateral identity. Peloric mutants derived from *cyc* result

from mutations in a second gene, *dichotoma* (*dich*), indicating that both *cyc* and *dich* are needed to establish full dorsoventral asymmetry.

Phenotype and development

Each mature wild-type flower has five sepals, five petals, four stamens and two united carpels. The petals are united for part of their length to form a corolla tube ending in five separate lobes. The petals can be classified according to size, shape and epidermal cell characteristics into three types: two dorsal, two lateral and one ventral. The dorsal petals have large lobes and are relatively free of hairs whereas the lateral and ventral petals have smaller lobes with yellow areas and hairs in the tube (Fig. 1). The ventral petal has bilateral symmetry whereas both the dorsal and lateral petals are individually asymmetric along the dorsoventral axis (this is more easily seen in the flattened petal lobe diagrams in Fig. 1). Alternating with the petals, three types of stamen primordia are initiated in whorl 3: one dorsal, two lateral and two ventral. The dorsal stamen primordium arrests at an early stage of development to form a small structure called a staminode; the two lateral stamens can be distinguished from the ventral pair by their shorter length and lack of hairs at their base. The filaments of the lateral and ventral stamens twist in a consistent way such that all anthers eventually face ventrally (floral diagram, Fig. 1). Looking down on the flower, filaments on the left twist clockwise whereas those on the right twist anticlockwise, implying that the lateral and ventral stamen filaments have individual asymmetry along the dorsoventral axis.

In *Antirrhinum* flowers with a peloric phenotype, there are typically six sepals, six petals, six stamens and two united carpels. All of the petals are bilaterally symmetrical and resemble the ventral petals of wild type, except that the lobes are slightly smaller (Fig. 1). The stamens resemble the wild-type ventral stamens and the filaments do not twist significantly so that all anthers tend to face towards the centre of flower (floral diagram, Fig. 1). The flowers therefore have a ventralized phenotype in which the number, type and internal symmetry of organs are affected.

The flowers of *cyc*-608 mutants usually have six sepals, six petals, six stamens and two united carpels. Most commonly, three petals have a ventral identity and the remaining petals have a combination of dorsal and lateral characteristics (Fig. 1). In some cases, however, the axis of asymmetry is differently aligned with respect to the organs, giving four petals with ventral identity and two with dorsal/lateral identity (Fig. 1, floral diagrams). Four or five stamens have a ventral identity, depending on the alignment of the dorsoventral axis, whereas the remaining stamens resemble the lateral stamens of wild type (Fig. 1). The twisting of stamen filaments is variable and the anthers end up facing in various directions (Fig. 1).

Development of wild type and mutant flowers was investigated by scanning electron microscopy (SEM). Floral meristems initiate sequentially at a rate of about one every 10 hours on the periphery of the inflorescence apex in the axils of bract primordia⁷. A sequence of developmental stages can be observed along the nodes of the inflorescence, starting with the earliest stage near the top (node 0), and progressively later stages below⁸. In wild type, the lateral and ventral sepal primordia first become visible at about node 10–11. The dorsal sepal primordium appears about a node later and its centre often seems to be offset to one side of the flower meristem (early stage 4, Fig. 2a). The dorsal primordium remains smaller than the other primordia through stages 4 and 5 (Fig. 2d, g). When petal and stamen primordia become clearly visible, their dorsal primordia are smaller than the others (stage 6, Fig. 2j). Eventually, development of the dorsal stamen arrests, whereas the growth of dorsal petals and sepals appears to catch up with the other organs in each whorl.

Differences between wild-type, semipeloric and peloric meristems were first detected at the time that sepal primordia started to appear. In peloric and semipeloric mutants, six sepal primordia usually emerge, although the two upper primordia are smaller, particularly in semipeloric meristems (Fig. 2a–f). By stages 5 and 6, six petal and six stamen primordia are visible, the upper petal and stamen primordia being noticeably smaller than the others in

semipeloric meristems (stage 6, Fig. 2k, l). Some peloric and semipeloric flowers develop with five instead of six organs in a whorl. The numbers of petals, sepals and stamens are often correlated with each other but there are some flowers with more sepals than petals or vice versa.

The *cyc* locus and mutant alleles

The *cyc*-608 allele arose from a transposon-mutagenesis experiment and is unstable, reverting to wild type at a rate of 0.34% (ref. 6). To determine whether *cyc*-608 was caused by a known transposon, genomic DNA from *cyc*-608 mutant and revertant plants was digested with various restriction enzymes and probed with different transposons. This showed that *cyc*-608 was caused by the insertion of a transposon belonging to the CACTA family^{9–12}, allowing the *cyc* locus to be cloned (Fig. 3). The sequence of the locus revealed an uninterrupted open reading frame (ORF) encoding a putative protein of 286 amino acids (Figs 3c, 4). RNA blots using a probe containing the ORF detected a transcript of 1.3 kb in the wild type but not in *cyc*-25 or *cyc*-608 mutants, and the transcript was found in young inflorescences but not in leaves (data not shown). A complementary DNA library was screened with a *cyc* probe and three cDNA clones containing sequence identical to the *cyc* genomic ORF were obtained. 5' and 3' RACE PCR (see Methods) confirmed the full length of the *cyc* transcript.

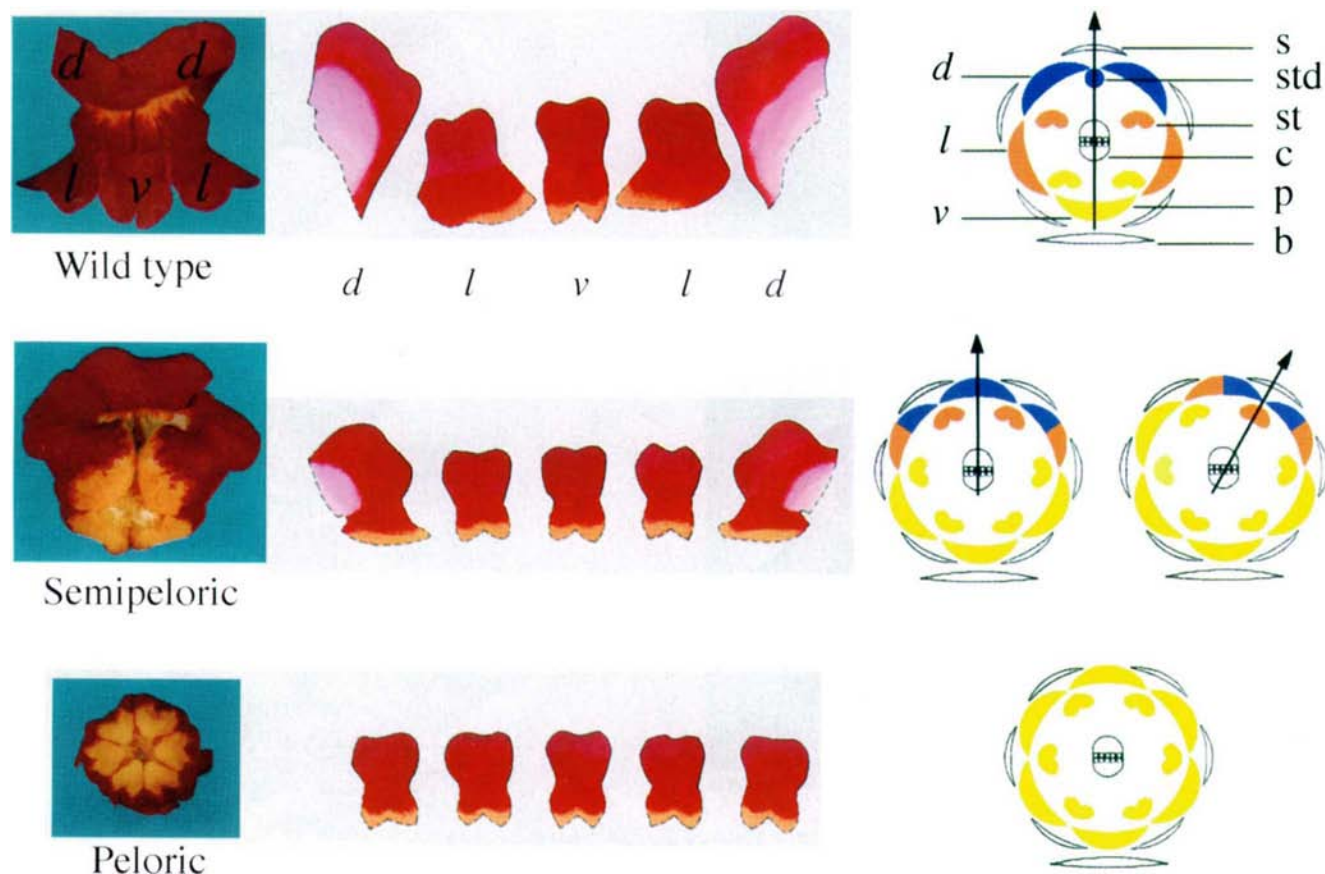


FIG. 1 Wild-type, semipeloric and peloric flowers. The wild-type *Antirrhinum* flower has an axis of dorsoventral asymmetry which is orientated such that the dorsal (upper or adaxial) part is nearer the stem whereas the ventral (lower or abaxial) part is nearer to the bract. On the left, flowers are photographed in face view with the dorsal (d), lateral (l) and ventral (v) petal lobes indicated for wild type. The characteristics of the different petal lobes are shown on the right of each flower, after they have been cut off the flower and flattened out (dotted lines indicate the sites of cutting). In the case of peloric and semipeloric flowers, only five petal lobes are shown for simplicity. The floral diagrams on the right show the relative position of different organs, with identities indicated by colours: dark blue (dorsal, top

diagram); brown (lateral); yellow (ventral). Semipeloric flowers lack petals with full dorsal identity but have petals regions with some dorsal characteristics (light blue, middle diagrams). The axis of asymmetry, as defined by the different types of organs is indicated by an arrow. In wild type, the axis of floral asymmetry is always aligned with the floral organs in a fixed way. For the semipeloric phenotype, the two floral diagrams indicate that the identity of floral organs varies depending on how the residual axis of asymmetry is aligned with respect to the organs. In peloric flowers there is no dorsoventral axis of asymmetry based on organ identity. Symbols: b, bract; c, carpel; p, petal; d, dorsal petal; l, lateral petal; v, ventral petal; s, sepal; st, stamen; std, staminode.

Based on restriction enzyme mapping and PCR, five *cyc* alleles were shown to have been generated by transposon insertions (Fig. 3c). Two of the alleles, *cyc*-650 and *cyc*^{neo}, carried insertions 5' to the ORF and had a slightly weaker phenotype than *cyc*-608 and *cyc*-25. The allele with the weakest phenotype, *cyc*^{abnor}, had an insertion in the second intron.

Analysis of the longest ORF of *cyc* shows that there is a bipartite motif, similar to a consensus nuclear localization signal (Fig. 4). This motif has been defined as two basic amino acids, followed by a spacer region, and five amino acids containing three basic residues¹³. No significant homology was found between *cyc* and other sequences in the gene banks, except for several expressed sequence tags (ESTs) from *Arabidopsis* with no known function.

The expression pattern of *cyc*

The expression pattern of *cyc* was analysed by RNA *in situ* hybridization. Digoxigenin-labelled antisense *cyc* RNA was used as a probe against inflorescence apices of wild type and *cyc* mutants. Serial sections through individual inflorescences were prepared to include different stages of floral development. In both longitudinal and transverse sections of wild type, *cyc* expression could only be detected in a dorsal region of developing floral meristems. Expression could be detected as early as at node 4, corresponding to stage 1 of floral development⁸. Expression of *cyc*

was initiated near the junction between inflorescence meristem and the floral meristem within a domain 2–4 cells wide, that gradually expanded as floral meristems grew larger (Fig. 5a, d). When the sepal primordia became visible, *cyc* expression was detectable in the dorsal sepal and in the dorsal part of the floral dome (Fig. 5b). In some wild-type genetic backgrounds, *cyc* expression was also detected in the dorsal part of the two lateral sepals (Fig. 5e). In late stage 4 (node 14), *cyc* expression was more concentrated in regions where the primordia of dorsal petals and the staminode were forming (Fig. 5e). When petal, stamen and carpel primordia were clearly visible (stage 6), *cyc* expression could only be detected within the two dorsal petals and the retarded staminode (Fig. 5c, f). In transverse sections, *cyc* expression in dorsal petals appeared to end abruptly a few cells away from the junction with the lateral petals (Fig. 5f). In more mature flowers, expression was weakly detected only in the staminode and the two dorsal petals (data not shown). Therefore, the asymmetric expression pattern of *cyc* was maintained during flower development. The inflorescences of different *cyc* mutants, that is, *cyc*-25 and *cyc*-608, did not show any expression of *cyc* (data not shown).

Analysis of peloric mutants

One explanation for the semipeloric phenotype of *cyc*-608 mutants was that other factors present in the genetic background were responsible for the residual dorsoventral asymmetry in the

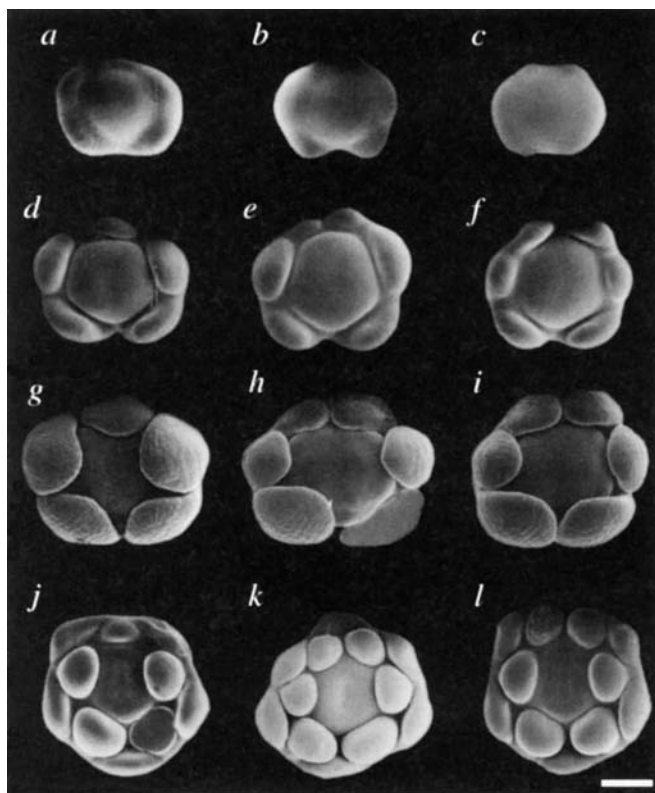


FIG. 2 Development of wild-type, semipeloric and peloric floral meristems as revealed by SEM. Four stages are shown: early stage 4 (sepal primordia initiate; top row, a–c), late stage 4 (sepal primordia are clearly visible; second row, d–f), stage 5 (petal primordia visible; third row, g–i) and stage 6 (sepals are removed to reveal all floral organ primordia; bottom row, j–l). In wild type, the dorsal sepal primordium appears about one node later than the lateral and ventral sepals (a), and is retarded (d) before catching up to be about the same size as the others by stage 5 (g). Similarly, the dorsal petal and dorsal stamen primordia appear later than the others, and are retarded (j). In semipeloric meristems, the dorsal sepal, petal and stamen primordia appear slightly later and are more retarded than those in other positions (b, e, h and k). In the peloric mutant derived from *cyc*-608, all the primordia of sepals, petals and stamens appear at about the same time, although a little retardation in the dorsal primordia is normally observed at initiation.

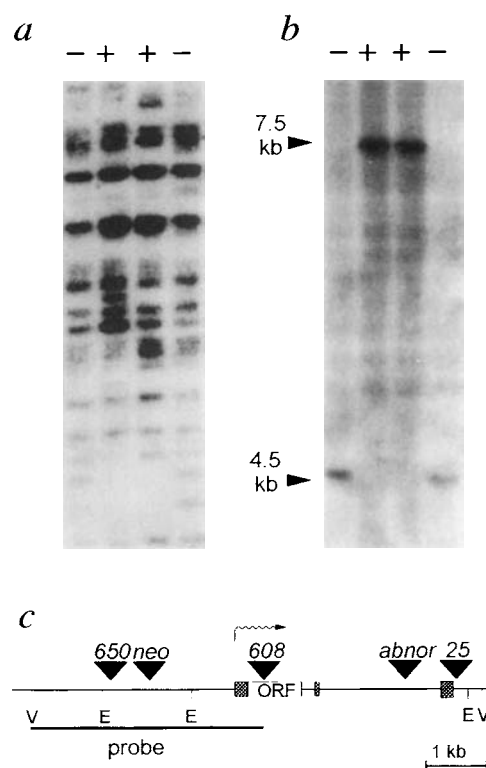


FIG. 3 Isolation and structure of the *cyc* gene. a, Genomic DNA from two *cyc*-608 mutant plants (–) and two independent revertants (+) were digested with *EcoRV* and blotted. The blot was probed with a 600-base pair (bp) fragment from the conserved end of Tam4 (refs 9–12), and a novel band of 4.5 kb was observed in mutants but not in revertants. b, The 4.5-kb fragment was cloned and the region flanking the transposon (c), was used to probe the same blot as in a, revealing the different banding patterns between mutants (4.5 kb) and revertants (7.5 kb). Different banding patterns between mutants and revertants were also observed in DNA from two other alleles *cyc*-25 and *cyc*-650, confirming that a fragment of the *cyc* locus had been cloned (not shown). c, Map of the *cyc* locus. The exons and predicted ORF are indicated in rectangles and the arrow indicates the direction of transcription. Black triangles represent the sites of transposon insertion for different *cyc* alleles. Restriction enzyme digestion sites within transposons are not shown. E, *EcoRI*; V, *EcoRV*.

flowers. This could be tested by analysing two peloric plants that had been derived from transposon-mutagenized *cyc*-608 plants. Neither of the peloric derivatives showed a detectable alteration within the 15-kilobase (kb) region of the *cyc* locus compared with their *cyc*-608 progenitor, indicating that the peloric phenotype might have resulted from a mutation in a second gene. To confirm this, one of the peloric mutants was backcrossed to wild type, and the phenotypes of F_2 plants were scored. In addition to peloric, semipeloric and wild type, there was a novel phenotype in which the dorsal petal lobes were separated from each other by a deeper divide. This was similar to the phenotype of *dichotoma* (*dich*) mutations^{14,15}, suggesting that the peloric mutants derived from *cyc*-608 had arisen by a mutation in the *dich* gene.

Analysis of an independently derived peloric line (stock-25) has shown that this also carries mutations in both *cyc* and *dich* (J. Almeida, M. Rocheta and L. Galego, personal communication). The *cyc*-608-derived peloric mutant was therefore crossed to stock-25 to test for complementation. All the plants in the F_1 and F_2 generations had peloric phenotypes, except for a few plants that displayed single mutant phenotypes, presumably as a result of either *cyc* or *dich* having reverted. To confirm further that the peloric phenotype resulted from the combination of *cyc* and *dich* mutations, the *cyc*-608 line was crossed to a line carrying *dich*. F_1 plants had wild-type flowers and the F_2 gave some plants with a peloric or semipeloric phenotype, as expected from segregation of two mutant alleles. However, the frequency of pelorics (13%) was higher than the expected 1/16, indicating that *dich* might not be behaving in a fully recessive manner. Further genetic analysis showed that *dich* had a variable phenotype and confirmed that it was not fully recessive in a *cyc* background (our unpublished results). This would account for the recovery of peloric plants in the direct progeny of *cyc*-608⁶, because only one copy of the *dich* gene needs to be inactivated to give peloric flowers in this background.

Discussion

The *cyc* and *dich* genes are both needed to set up full dorsoventral asymmetry in *Antirrhinum* flowers. Mutants that lack both *cyc* and *dich* activity have radially symmetrical peloric flowers, mutants for *cyc* alone have semipeloric flowers and *dich* single mutants have flowers with altered dorsal petals. This implies that the *cyc* and

dich genes can substitute for each other to some extent, allowing for residual asymmetry in each single mutant. One possibility is that *cyc* and *dich* encode related products that function in a similar way. We have recently confirmed this by isolating *dich* and showing that its sequence and expression pattern are indeed very similar to those of *cyc* (our unpublished results). However, the role of *cyc* appears to be more critical than *dich* for establishing dorsoventrality as its mutant phenotype is more extreme.

The *cyc* gene encodes a protein with no significant homology to other known proteins, although it does contain a putative nuclear localisation signal, indicating that it may play a role in transcriptional regulation. Transcripts of *cyc* accumulate specifically in the dorsal region of wild type floral meristems, consistent with its role in setting up dorsoventral asymmetry. Its expression is first detected at a very early stage, before any morphological dorsoventral asymmetry in the meristem is visible by SEM, and continues through to later stages in the dorsal petal and stamen primordia (Fig. 6a–e). The phenotypic consequences of this expression pattern can be conveniently divided into early effects on primordium initiation and later effects on organ morphology.

Early effects of *cyc* on primordium initiation. During early stages, *Cyc*⁺ activity retards growth and reduces the number of primordia in dorsal regions of the flower meristem. In meristems of semipeloric flowers, two dorsal sepal primordia are typically formed and their development is almost in synchrony with the other primordia (Fig. 6f–j). In contrast, wild-type floral meristems form a single dorsal sepal primordium which is retarded in growth relative to the other sepal primordia. The region of retarded growth corresponds to the domain where *cyc* transcripts accumulated in wild type, indicating that localized *Cyc*⁺ activity represses growth and/or primordium initiation (Fig. 6). It is possible that two dorsal primordia are initiated in wild type but that only one continues to develop and become visible. This might account for the observation that the dorsal sepal primordium often appears to be displaced to one side of wild-type meristems in early stages of growth (Fig. 6d, e). Increased sepal number in *perianthia* mutants of *Arabidopsis* also correlates with more synchronous sepal initiation, although in this case the wild-type gene appears to advance growth in ventral regions rather than retard it in dorsal regions¹⁶. Expression of *cyc* in emerging dorsal petal and stamen primordia correlates with their reduced growth and number, indicating that *cyc* may affect all dorsal organ primordia in a similar way at early stages (Fig. 6e). Unlike *cyc*, the *dich* gene does not appear to have a major effect on organ number.

The early effects of *cyc* may help to bring the sites of organ initiation into register with the axis of asymmetry. By reducing growth, *Cyc*⁺ activity could ensure that the sites of dorsal organ primordia become centred within the *cyc* expression domain. In this way, asymmetry could be coupled more consistently with the position of primordia in the flower. A prediction is that in *cyc* mutants, the residual asymmetry conferred by *Dich*⁺ activity should be more variably aligned with respect to organ primordia. This appears to be the case, because some *cyc* mutant flowers have two dorsal/lateral and four ventral petals rather than the more usual three dorsal/lateral and three ventral petals (see floral diagrams in Fig. 1).

Later effects of *cyc* on organ morphology. Transcripts of *cyc* are detected in dorsal petal and staminode primordia, from their inception through to later stages of development. The domain of *cyc* expression ends at a boundary lying a few cells away from the junction between dorsal and lateral petals in whorl 2, and around the staminode in whorl 3 (Fig. 5f). The effects of *cyc* expression on the developmental fate of dorsal primordia vary between whorls. In whorl 2, the dorsal petal lobes of wild type are larger in area than in *cyc* mutants, indicating that a role of *Cyc*⁺ activity is to promote petal lobe growth. In contrast, in whorl 3, the dorsal stamen of wild type becomes increasingly retarded relative to the other stamens, indicating that *cyc* activity in the dorsal stamen primordium acts to slow down and eventually arrest growth. The later effects of *cyc* on organ development therefore vary according

```

1 ATGTTGGGAAGAACACATACCTACATCTCCCTCAGGTTTCATCTCCCTCCTCACTCTCGGCGCGTACT 69
1 M F G K N T Y L H L P Q V S S S L H S R A A C T 23
70 TCTGTTGCTGATCTCAACGGCAATGAGATCCAGCTCCAGACGATGCTCTCGGCTTACTTAACACAC 138
24 S V V D L N G N E I Q L H D M L S G H Y L T T 46
139 GCGAATGCTCGGTTCTTGATCCACCGCTTGTTCACACACACACACATCAATCAGGCTGCTT 207
47 A N A F V L E S T A L F N N N N N F N H D V V 69
208 AATGGCTAAACAGAGATCTCTCTCCACATTTCCGACAAAACAGGCGGTGAAGAGATCGCCACAGC 276
77 N G L N R D P S S T F P T K Q A V K K D R H S 92
277 AAAATATACATCACAAGGCCAAGGACAGGAGCTCGGTTTATCCATCGGATTCGCAAGAGTTC 345
93 K I Y T S Q G P R D B R V R L S I G I A R K F 115
346 TTTGCTTACAGAGATGCTAGGTTTCGACAGCGAGCAAAACCTTGATTGGCTGCTCAATAGTGC 414
116 F D L Q E M L G F D K P S K T L D W L L T K S 138
415 AAAACCGCTATCAAGAGCTTGTGCGTCTAAAAGTACTAAGAGCACTCTCTCCCTTGTGATGAT 483
139 K T A I K E L V Q S K S T K S N S S P C D D 161
484 TGTGAGGAAGTTGTGTCTGTAGATCTTGAAGATGACAGATCAITCAAGGGGGAATCACTGAAGCT 552
162 C E E V V S V D S E N V T D H S K G K S L K A R 184
553 AACACAAATGTAAGAGGAATGGATTCACATCAAGCAGCAGCAAGAGATCGAGGCGAAGCGCAGGA 621
185 N N K C K E A M D S H Q A A A K E S R A K A R 207
622 GCGAGAGCTAGGGAAGAACTAAGGAGAAGTGTGTATCAAACTGAATGAAGCTATAGTACTGAGG 690
208 A R A R E R T K E K M C I K Q L N E A I V L R 230
691 AACCATCAATTGAAGTTCCGGTACCGGTGAGGCGTTTGTTCATCACTGTTCCGGTTTCATGAGCA 759
231 N H Q F E V S G T R E A F V H P V F G F H Q Q 253
760 AACTATGGCAACGCATCTCAGAACTGGGATCAGAGTAACCTTCTCGCACTCGAAGCAGCTATSC 828
254 N Y G N A S H E N W D Q S N J S S Q S N Q L C 276
829 GCCATTTTGAATCAGCACAAGTTCATCAATTAG 861
277 A I L N Q H K F J N * 286

```

FIG. 4 Sequence of the *cyc* ORF and the deduced protein sequence. The underlined amino acids indicate the putative bipartite motif similar to a consensus nuclear localization signal¹³. The triangle indicates the insertion site of the transposon in *cyc*-608. A few ESTs (accession numbers: T45419, R29994 and R30409) from *Arabidopsis* have been found to share some homology with *cyc*.

to the whorl, in contrast to the early effects on growth rate and organ number which are more similar between whorls. This may be because early *cyc* expression occurs before that of organ identity genes (that is, before stage 4; refs 17, 18), whereas at later stages, *cyc* can interact in a combinatorial fashion with organ identity genes.

Even though *cyc* transcripts cannot be detected outside the dorsal domain, *cyc* has clear effects on the development of lateral organs, altering petal shape and stamen length. We cannot rule out the possibility that *cyc* transcripts are present in lateral organs below the level of detection. However, it seems more likely that *cyc* expression in the dorsal domain acts non-autonomously, through cell–cell signalling, to influence the behaviour of lateral organ primordia. Perhaps the flower meristem is first divided into two main regions: a dorsal domain expressing *cyc* and the remaining domain with no *cyc* expression. This early partition could then be further elaborated by cell–cell interactions to generate three distinct domains: dorsal, lateral and ventral. Candidate genes involved in this further elaboration are the *radialis* gene⁶, which gives a phenotype similar to *cyc* and *divaricata*, which confers a lateralized phenotype (ref. 15; J. Almeida, M. Rocheta and L. Galego, personal communication).

In addition to establishing distinctions between dorsal, lateral and ventral organs, the *cyc* and *dich* genes are also required to set up dorsoventral asymmetry within individual organs. Unlike the

ventral petal, which is bilaterally symmetrical, dorsal and lateral petals are asymmetric along the dorsoventral axis of wild-type flowers. Similarly, the left- or right-handed twists of lateral and ventral stamen filaments indicates that they may have different growth rates along the dorsoventral axis, ensuring that the anthers are presented in specific orientations. These aspects of internal petal and stamen asymmetry are lost in *cyc:dich* double mutants, suggesting that in addition to setting up differences between organs, *cyc* and *dich* are also needed to establish subdomains within organs along the dorsoventral axis.

The *cyc* expression pattern, together with its phenotypic effects, raise the question of how its role has evolved. It is possible that the early effects of *cyc* on retarding growth rate and primordium initiation are more highly conserved than the later effects on organ morphology. For example, in many species belonging to the same family as *Antirrhinum* or to closely related families, the dorsal petal lobes are smaller than the lateral or ventral lobes, indicating that in these cases *cyc* homologues may act to reduce final lobe size rather than increase it as in *Antirrhinum*. It has also been proposed that early aspects of floral asymmetry in the Leguminosae are more conserved than later characteristics¹⁹, although it is not clear if *cyc* homologues are involved in this case because dorsoventral asymmetry in this family is thought to have evolved independently from that in *Antirrhinum*. It is also unclear whether any of the early or late roles of *cyc* might also be

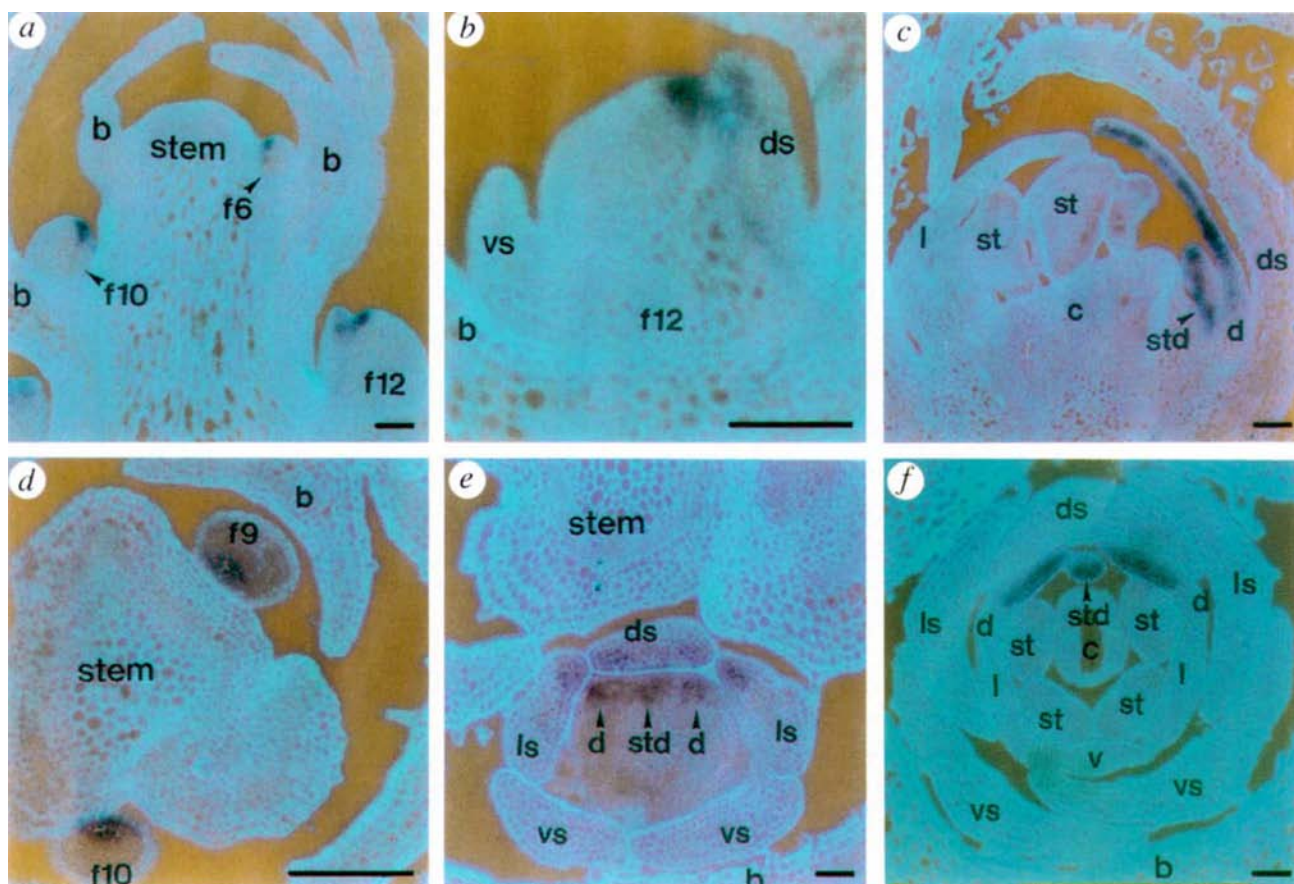


FIG. 5 RNA *in situ* hybridization of wild-type sections probed with *cyc*. The top row (a–c) shows longitudinal sections and the bottom row (d–f) shows transverse sections. The sections were probed with digoxigenin-labelled *cyc* antisense RNA and viewed under bright field, which gives a dark blue signal on turquoise background tissue. a, Section through an inflorescence apex; the node of each floral primordium is indicated. The signal can be seen in the dorsal (upper) regions of each floral meristem. b, Sections through a node-12 floral meristem, showing a signal in the dorsal region of the floral dome and in the dorsal sepal primordium. c, Floral meristem at a late

stage, showing the signal in the dorsal petal and staminode. d, Floral primordia at stage 2, showing a signal in the dorsal regions near the junction with the inflorescence stem. e, Stage-5 meristem, showing the signal in the primordia of the dorsal sepal, part of the lateral sepals, dorsal petals and staminode. f, Floral meristem at a late stage, showing the signal in the staminode and dorsal petals. Symbols: b, bract; c, carpel; p, petal; d, dorsal petal; l, lateral petal; v, ventral petal; s, sepal; st, stamen; std, staminode; f, floral meristem (node number indicated); ds, dorsal sepal; ls, lateral sepal; vs, ventral sepal. Scale bar, 100 μ m.

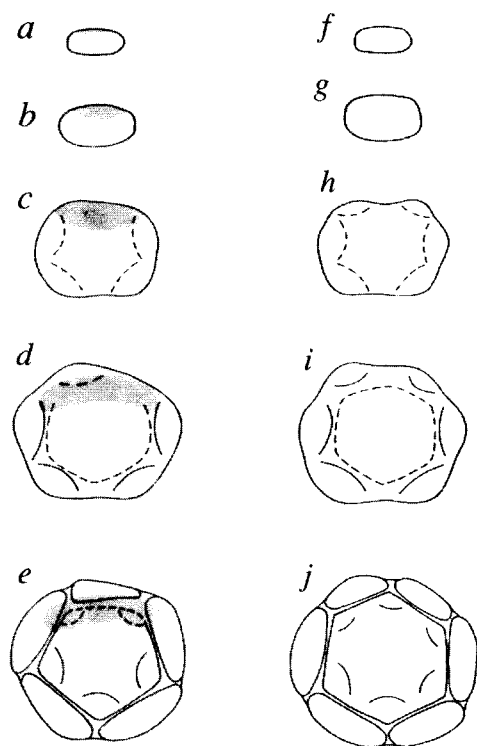


FIG. 6 Summary of the *cyc* expression pattern and its early effects on floral development in wild type as compared to the *cyc*-608 mutant. a–e, Wild type, showing domain of *cyc* expression (grey area) at various developmental stages based on reconstructions from serial sections. f–j, Semipeloric *cyc*-608 mutant, showing more growth and primordium initiation in the region where *cyc* is normally expressed in wild type. Solid lines indicate outlines of clearly visible primordia; dotted lines indicate regions of primordium initiation.

present under the guise of a different expression pattern in species with radially symmetrical flowers. The isolation and analysis of *cyc* homologues from species with different patterns of floral symmetry should allow some of these possibilities to be investigated. □

Methods

Plant Material. Plants of wild type (JI-75), semipeloric (JI-608) and a peloric mutant derived from *cyc*-608 (JI-659), were grown either in the greenhouse or the field as described previously²⁰. Lobes were dissected from flowers and then flattened between two glass slides before drawing. For floral diagrams, five flowers from six plants of each phenotype grown in the greenhouse were examined. For SEMs, plants were grown in growth rooms at 20 °C. SEMs were made on plastic replicas as described⁸.

DNA and RNA analysis. DNA extraction and blot analysis were done as described previously¹². A 600-bp DNA from the 3' end of Tam4, conserved among the CACTA transposon family in *Antirrhinum majus*^{9–12}, was used as a probe to reveal a 4.5-kb *EcoRV* band which was only present in the mutants. The 4.5-kb *EcoRV* fragment was cloned as pJAM608, using the λ gt10 vector (Amersham PRN1713, N334L). The DNA flanking the transposon in pJAM608 was used to screen a genomic library (kindly provided by H. Sommer) to obtain λ JAM151 containing the *cyc* locus. Most of the 15-kb insertion in λ JAM151 was sequenced and the ORF identified. Restriction enzyme site mapping was used to identify the approximate insertion positions of the transposons in different *cyc* alleles. Four alleles, *cyc*-25 and *cyc*-608 from the John Innes collection; *cyc*^{abnormis} (*cyc*^{abnorm}) and *cyc*^{neohemiradialis} (*cyc*^{neoy}) from the Gatersleben collection, have been described^{6,14,15,21}. The *cyc*-650 allele was obtained in a transposon-mutagenesis experiment (R.C. and E.C., unpublished results). PCR on genomic DNA of each allele was done to determine the exact transposon insertion site, using oligonucleotides to the conserved end of the CACTA transposon family and to the *cyc* sequence. DNA spanning the ORF was used as a probe to screen a cDNA library made from young inflorescence of wild type²², and from about 2×10^6 recombinants, three independent cDNA clones (pJAM167, pJAM168 and pJAM169) were obtained and sequenced. Two of them had the same structure, indicating that the *cyc* gene had three exons, the ORF being contained within the first exon. The third cDNA was a 2.1-kb variant, containing an unspliced second intron. The transcription start and full-length cDNA were confirmed by 3' and 5' RACE PCR (5' RACE system: GibcoBRL).

The methods for digoxigenin labelling of RNA probes, tissue preparation and *in situ* hybridization were done as described²³. The poly(A)⁺ tail of a *cyc* cDNA subclone, pJAM167, was deleted to obtain a plasmid pJAM193, which was used to generate antisense and sense probes using either T3 or T7 polymerases.

Received 1 August; accepted 26 September 1996.

- Coen, E. S. & Nugent, J. M. *Development* (suppl.) 107–116 (1994).
- Coen, E. S. & Meyerowitz, E. M. *Nature* **353**, 31–37 (1991).
- Weigel, D. & Meyerowitz, E. M. *Cell* **78**, 203–209 (1994).
- Coen, E. S. *Annu. Rev. Pl. Physiol. Pl. Mol. Biol.* **42**, 241–279 (1991).
- Linnaeus, C. *De Peloria*. Diss. (Amoenitates Acad., Uppsala, 1749).
- Carpenter, R. & Coen, E. S. *Genes Dev.* **4**, 1483–1493 (1990).
- Vincent, C., Carpenter, R. & Coen, E. S. *Curr. Biol.* **5**, 1449–1458 (1995).
- Carpenter, R. et al. *Pl. Cell* **7**, 2001–2011 (1995).
- Bonas, U., Sommer, H. & Saedler, H. *EMBO J.* **3**, 1015–1019 (1984).
- Krebbers, E. et al. *Mol. Gen. Genet.* **209**, 499–507 (1987).
- Somer, H. et al. in *Plant Transposable Elements* (ed. Nelson, O. E.) 227–235 (Plenum, New York, 1988).
- Luo, D., Coen, E. S., Doyle, S. & Carpenter, R. *Plant J.* **1**, 59–69 (1991).
- Dingwall, C. & Laskey, R. A. *Trends Biochem. Sci.* **16**, 478–481 (1991).
- Kuckuck, H. & Schick, R. Z. f. indukt. Abst.-u. Vererbungsl. **56**, 51–83 (1930).

- Stubbe, H. *Genetik und Zytologie von Antirrhinum L. sect. Antirrhinum* (Frischer, Jena, 1966).
- Runnig, M. P. & Meyerowitz, E. M. *Development* **122**, 1261–1269 (1996).
- Sommer, H. et al. *EMBO J.* **9**, 605–613 (1990).
- Bradley, D., Carpenter, R., Sommer, J., Hartley, N. & Coen, E. S. *Cell* **72**, 85–95 (1993).
- Tucker, S. in *Contemporary Problems in Plant Anatomy* (eds White, R. A. & Kickison, W. C.) 351–395 (Academic, London, 1984).
- Carpenter, R., Martin, C. & Coen, E. S. *Mol. Gen. Genet.* **207**, 82–89 (1987).
- Hammer, K., Knüpfner, H. *Kulturpflanze* **38**, 91–117 (1990).
- Simon, R., Carpenter, R., Doyle, S. & Coen, E. S. *Cell* **78**, 99–107 (1994).
- Coen, E. et al. *Cell* **63**, 1311–1322 (1990).

ACKNOWLEDGEMENTS. We thank N. Hartley for sequencing of the genomic region of *cyc*; P. Walker for care and maintenance of plants; J. Almeida for helpful discussions; D. Bradley, P. Cubas, M. Chadwick, G. Ingram, Y. Xue and S. Doyle for critical reading and comments on the manuscript. This work was supported by the Gatsby Foundation and BBSRC PMB2 initiative.

CORRESPONDENCE and requests for materials should be addressed to E.C.

Origin of the magnetic spiral arms in the galaxy NGC6946

Zuhui Fan & Yu-Qing Lou

Department of Astronomy and Astrophysics, The University of Chicago, Chicago, Illinois 60637, USA

THE recent discovery¹ of coherent, large-scale magnetic spiral arms located between the optical spiral arms of the nearby galaxy NGC6946 is difficult to reconcile with current galactic dynamo theory². The formation of optical spiral arms is generally attributed to the role of density waves in the gas of the galactic disk, which trigger the formation of massive, luminous stars; turbulent processes in the gas are also believed to play a key role in the generation of the galactic magnetic field. Here we present the results of magnetohydrodynamical (MHD) calculations, which show that both fast ($\sim 20 \text{ km s}^{-1}$) and slow ($\sim 0.2 \text{ km s}^{-1}$) MHD density waves can exist in a rotating galactic disk. Slow MHD density waves occur over the portion of a galactic disk that rotates almost rigidly³, and can give rise to a pattern of polarized radio emission similar to that observed in NGC6946, where the magnetic field is strong between the optical spiral arms. In contrast, fast MHD density waves produce spiral magnetic structures that are roughly coincident with the optical spiral arms—and which extend into the differentially rotating disk—as observed in the Whirlpool galaxy, NGC5194 (refs 4–6).

Nearly face-on, with an inclination angle of $\sim 30^\circ$ to the sky plane, NGC6946 ($\sim 5\text{--}10 \text{ Mpc}$ away) has no apparent companion galaxy and may^{7,8} or may not⁹ have a weak central bar. Both of its optical and magnetic spiral arms occupy a disk angular radius of $\sim 4 \text{ arcmin}$ (refs 1, 3) within which the rotation speed V_θ first increases linearly and then slowly bends to approach a constant value with increasing radius^{10–12} (we shall refer to this situation as almost or nearly rigid rotation); in the radial range from ~ 4 to 11 arcmin , $V_\theta \approx 200 \text{ km s}^{-1}$ remains roughly constant^{10–12}. In contrast, total magnetic and gaseous spiral structures nearly overlap in M51^{3–6} (NGC5194), although the actual ridge lines of strongest radio synchrotron emission lie along the dark dust lanes slightly inside the optical arms that are outlined by giant bright H II complexes; this slight shift is attributed to the time delay^{5,13} of $\sim 10^7 \text{ yr}$ in star formation as the peaks of density waves^{14–16} sweep by. Both the optical and magnetic spiral structures of M51 extend far into the disk portion with an almost constant $V_\theta \approx 200 \text{ km s}^{-1}$. Except for several optical/radio inter-arm links and distortions caused by the companion NGC5195, the ‘grand design’ of M51 is a prime example showing the correlation between gas and total magnetic field^{5,6,17} in gas-rich spiral galaxies. By comparison, the correlation between gas arms and ridges of polarized radio emission in M51 is not as close.

Despite some irregularities and fuzziness in the optical image, the multi-armed spirals of NGC6946 can be readily recognized in observations at several wavelengths. After examining the radio maps of NGC6946 (ref. 1), we were prompted to seek a global mechanism for generating such large-scale coherent magnetic spiral arms interlaced with optical spiral arms. It is worth noting that NGC6946 is not alone in having this type of structure; gas-rich late-type galaxies IC342¹⁸ and M83 (NGC5236)¹⁹ also have somewhat similar magnetic arms that may be observed in polarized radio emissions. Galactic spiral structures in both IC342 and M83 also occupy the almost rigidly rotating disk portions³. Thus in a broader perspective, a theory for galactic density waves incorporating magnetic-field effects should accommodate both types of spiral structures for density and magnetic field as shown by NGC6946 and M51. With these considerations in mind, we will show here that both fast and slow MHD density waves can exist

under appropriate galactic conditions. For typical nearby spiral galaxies, the Alfvén and sound speeds are comparable and fall in the range $\sim 10\text{--}20 \text{ km s}^{-1}$. In general, MHD density waves might be excited by a central bar, a satellite galaxy, or by an intrinsic amplification and feedback mechanism^{13,20,21}.

For conceptual clarity, we first describe a familiar case of compressible MHD waves in a uniformly magnetized medium. There exist fast and slow MHD waves whose dispersion relations can be readily derived. Fast MHD waves can propagate in all directions, whereas slow MHD waves cannot propagate perpendicular to the background magnetic field. For a weak magnetic field, slow MHD waves are more magnetic in nature and gradually disappear in the limit of zero magnetic field, whereas fast MHD waves are more acoustic in nature. The perturbations of parallel magnetic field and density are positively and negatively correlated for fast and slow MHD waves, respectively. For slow MHD waves with relatively small parallel wavenumbers, if the Alfvén speed C_A is much higher or lower than the sound speed C_s , the magnitude of relative density fluctuation will be respectively much larger or smaller than the magnitude of relative parallel magnetic field fluctuation; these two relative fluctuations are comparable in magnitude when $C_A \approx C_s$.

As analogues of the above fast and slow MHD waves, fast and slow MHD galactic density waves exist in an axisymmetric thin rotating gaseous disk with $B_\theta \propto r^{-1}$, where the cylindrical coordinates (r, θ, z) are used. One can readily write out MHD equations for perturbations of magnetic field $\mathbf{b} \equiv (b_r, b_\theta, 0)$, velocity $\mathbf{v} \equiv (v_r, v_\theta, 0)$, pressure p , surface mass density μ and the negative gravitational potential ϕ . The $\exp(i\omega t - im\theta)$ dependence is assumed for these variables, where ω is the angular frequency in an inertial reference frame and the integer $m \geq 0$ gives the number of spiral arms. The mathematical procedure for analysing these MHD density-wave equations can be succinctly summarized as a combination of Binney and Tremaine’s approach¹³ plus MHD. As in the case of hydrodynamic density waves^{13–16,22}, $\phi(r, \theta, 0)$ and $\mu(r, \theta)$ are connected through an integral relationship by solving Poisson’s equation. By invoking the well-known WKBJ approximation^{14,15,23} with a sufficiently large radial wavenumber k , ϕ and μ are related by

$$\frac{1}{r^{1/2}} \frac{\partial(r^{1/2}\phi)}{\partial r} \Big|_{z=0} \approx 2\pi G i \operatorname{sgn}(k) \mu + \mathcal{O}\left[\frac{\mu}{(kr)^2}\right] \quad (1)$$

where the last term represents a fractional error in the order of $(kr)^{-2}$, where G is the gravitational constant and $\operatorname{sgn}(k) = \pm 1$ for $k \gtrless 0$, corresponding to leading/trailing arms. Consistent with equation (1), azimuthal derivatives of perturbations are retained. By examining all relevant terms to the leading order of large kr , we derive the local dispersion relation

$$(\omega - m\Omega)^2 \approx \kappa^2 + k^2(C_A^2 + C_s^2 - 2\pi G \mu_0 / |k|) \quad (2)$$

for fast MHD density waves, where $\Omega \equiv V_\theta/r$ is the disk angular rotation rate, $\kappa^2 \equiv (2\Omega/r)d(rV_\theta)/dr$ defines the epicyclic frequency κ , and μ_0 is the disk surface mass density. It may readily be shown that μ and b_θ are roughly in phase to leading order of large kr ; this is true for either a rigidly rotating disk or a disk with a constant V_θ . With a positive right-hand side, equation (2) generalizes Toomre’s stability criterion²⁴ for a non-magnetized thin rotating disk which can be locally stabilized by its rotation and sound speed. Equation (2) may be solved to obtain two values of $|k|$. The short-wave (larger $|k|$) solution is naturally consistent with the WKBJ approximation; although heuristic, the long-wave (smaller $|k|$) solution must be complemented by numerical investigations^{13,21}. The radial group velocity²⁴ for a packet of fast MHD density waves can be derived from equation (2); the propagation direction is towards the co-rotation point r_c on the long-wave branch and away from r_c on the short-wave branch. Apparently, the presence of B_θ tends to enhance the local stability

of fast MHD density waves. The numerical MHD model in the co-rotating pattern frame of reference developed by Roberts and Yuan¹⁷ is essentially for nonlinear fast MHD density waves which can steepen into shocks. Radio observations^{5,6} of M51 confirmed this model. The presence of considerable random magnetic fields in the total-radio-emission arms can be attributed to various activities usually associated with high-density spiral arms.

Using the WKB approximation (equation (1)) for MHD perturbations in a self-gravitating thin rigidly rotating disk (for example, a Maclaurin disk¹³), we derive

$$(\omega - m\Omega)^2 \approx \frac{k^2 C_A^2 (C_s^2 - 2\pi G \mu_0 / |k|) m^2 / r^2}{\kappa^2 + k^2 (C_A^2 + C_s^2 - 2\pi G \mu_0 / |k|)} \quad (3)$$

for slow MHD density waves²⁵⁻²⁷ to leading order of large kr , where $C_A \approx C_s$ and $\Omega / C_A \approx k$ are used. From equation (3) and relevant MHD equations, it may be demonstrated that, to leading order of large kr , μ leads or lags b_θ more than $\pi/2$ in phase for trailing waves with the wave pattern speed $\omega_p \equiv \omega/m < \Omega$ or $\omega_p > \Omega$, respectively. We recall that for a uniformly magnetized medium in an infinite space without rotation, magnetic pressure perturbation anti-correlates (that is, has a phase difference of π) with density perturbation for slow MHD waves. Here, it is the rapid disk rotation that modifies the phase difference between μ and b_θ for slow MHD density waves. Equation (3) may be solved for $|k|$; only one root $|k| > 0$ is physical, as the other negative root must be discarded. For a packet of slow MHD density waves, the radial group velocity C_G derived from equation (3) is given by

$$C_G \equiv \frac{\partial \omega}{\partial k} \approx \frac{k C_A^2 [\kappa^2 (C_s^2 - \pi G \mu_0 / |k|) + C_A^2 \pi G \mu_0 |k|] m^2 / r^2}{(\omega - m\Omega) [\kappa^2 + k^2 (C_A^2 + C_s^2 - 2\pi G \mu_0 / |k|)]^2} \quad (4)$$

with $C_G \geq 0$ for inward/outward propagations, respectively. In the linear treatment, global patterns of leading and trailing slow MHD density waves with ω_p greater or less than the nearly constant Ω could all exist; other processes must be explored to determine the pattern selection conditions. For trailing ($k < 0$) waves with $\omega_p < \Omega$, a packet of slow MHD density waves moves inwards. For both fast and slow MHD density waves, the direction of magnetic field perturbation is largely aligned with the magnetic spiral arm to leading order of large kr .

It is clear from equation (3) that slow MHD density waves disappear for $m = 0$ or $B_\theta = 0$ as expected, and their phase speeds are less than either C_A or C_s . The local stability of slow MHD density waves requires $C_s^2 > 2\pi G \mu_0 / |k|$ as if rotation and magnetic field were absent²⁵⁻²⁷; this stability criterion appears more stringent than that of Toomre's^{13,24}. Slow MHD density waves with ω_p close to Ω can appear in a thin disk portion with a nearly constant Ω until a radius beyond which a sufficiently strong differential rotation prevails. By equation (3), the higher the values of C_s and C_A , the larger the tolerance for deviations away from a constant Ω . Thus, slow MHD density waves can also appear in the part of the disk which has a sufficiently weak differential rotation; for example, in the transition zone between the rigid rotation and the flat part of the rotation curve. If equation (3) was applied to a disk portion with a constant V_θ , slow MHD density waves could only appear within a relatively narrow strip around the co-rotation radius r_C .

Both CO and H I observations¹⁰⁻¹² indicate that NGC6946 rotates almost rigidly up to an angular radius of ~ 4 arcmin. Multi-armed ($m = 4$) optical and magnetic spirals⁹ appear roughly within that angular-radius range^{1,3}. Large-scale coherent magnetic arms revealed by polarized radio emission¹ lie between the optical arms with a radial magnetic-optical inter-arm distance closer to the outer optical arm. For neutral (that is, neither growing nor decaying) slow MHD density waves to appear, it is necessary that $C_s^2 > 2\pi G \mu_0 / |k|$ in equation (3). This would have been difficult for not so large a $|k|$ as shown by the radio and optical maps^{1,3} and a $\mu_0 \lesssim 10^{-2} \text{ g cm}^{-2}$, because the oft-quoted value for sound speed C_s is only $\sim 10 \text{ km s}^{-1}$. However, if one takes

into account possible turbulent gas and molecular cloud velocity dispersions, as well as a relatively high stellar velocity dispersion (for example $\sim 30 \text{ km s}^{-1}$), C_s may be replaced by an effective velocity dispersion σ_N that is several times larger than C_s , and then the inequality $\sigma_N^2 > 2\pi G \mu_0 / |k|$ could be met. With these qualifications, we propose that the global pattern of optical-magnetic spiral structures in NGC6946 could be a manifestation of slow MHD density waves in an almost rigidly rotating portion of a thin galactic disk. The fact that the magnetic spiral arms in NGC6946, as revealed by polarized radio emissions, appear relatively clean and regular can be explained by reduced activity of the sort usually associated with relatively high-density spiral arms. As the magnetic spiral arm is somewhat closer to the outer optical arm¹ and $|k|$ appears to gradually decrease with increasing r , we infer that $\omega_p < \Omega$ for slow MHD density waves in NGC6946 (following the conventional wisdom that the spiral arms are probably trailing).

We note several caveats, as follows. First, it is known that the thin gaseous disk model¹³⁻¹⁶ presented here is, in the absence of B_θ , subject to non-axisymmetric global instabilities^{21,22} such as bar instabilities^{28,29} even if Toomre's stability criterion²⁴ is met. An azimuthal magnetic field cannot fundamentally alter this situation, because magnetic energy is a small fraction of galactic rotation energy. As before, one must invoke either a sufficiently large gas and stellar velocity dispersion^{22,30} or a massive dark matter halo²⁹ to prevent global instabilities in a thin magnetized galactic disk. Second, magnetized gas flows in spiral galaxies can lead to nonlinear MHD shock development. In the simple case of a uniformly magnetized medium, the density jumps across both fast and slow MHD shocks, whereas the tangential magnetic field increases and decreases across quasi-perpendicular fast and slow MHD shocks, respectively. Although the disk differential rotation may introduce complications, qualitatively similar phase relationships are expected for density and tangential magnetic field as in the cases of fast and slow MHD density waves. It is also possible that fast and slow MHD density waves coexist in the almost rigidly rotating portion of a spiral galaxy, although the present theory cannot determine their respective strengths nor their nonlinear couplings. Last, we note that we started this analysis by assuming the presence of an azimuthal magnetic field *a priori*, be it primordial or galactic-dynamo-generated. This bypasses the fundamental question as to the origin of galactic magnetic fields^{31,32}. All extant galactic dynamo theories² cannot explain the significant phase shift between magnetic and optical spiral arms seen in NGC6946¹. A combination of galactic-dynamo theory² and MHD density wave theory seems necessary³³, although several recent analyses³⁴ ignored the feedback of Lorentz force on flows and thus excluded the possibility of fast and slow MHD density waves.

Late-type gas-rich spiral galaxies IC342¹⁸ and M83¹⁹ share several features of NGC6946¹ and show polarized radio emissions outside optical spiral structures, although differences exist and present-day resolution in radio observations is not sufficiently high to resolve detailed magnetic structures. Because the radii of optical spiral structures in both IC342 and M83 are roughly comparable to their corresponding radii of almost rigid rotations³, these galaxies are possible candidates for displaying spiral patterns of slow MHD density waves. In fact, a systematic polarized radio search for spiral galaxies with probable slow MHD density waves is extremely desirable, and should be performed on all known gas-rich spiral galaxies whose radii of optical spiral structures are less than (or comparable to) their corresponding radii of almost-rigid rotations^{3,13}. $\square$

Received 4 June; accepted 24 September 1996.

1. Beck, R. & Hoernes, P. *Nature* **379**, 47-49 (1996).
2. Ruzmaikin, A., Sokoloff, D. & Shukurov, A. *Nature* **336**, 341-347 (1988).
3. Kormendy, J. & Norman, C. A. *Astrophys. J.* **233**, 539-552 (1979).
4. Elmegreen, B. G., Elmegreen, D. M. & Seiden, P. *Astrophys. J.* **343**, 602-607 (1989).
5. Mathewson, D. S., van der Kruit, P. C. & Brouw, W. N. *Astron. Astrophys.* **17**, 468-486 (1972).
6. Neinger, N. & Horellou, C. in *Polarimetry of the Interstellar Medium* (eds Roberge, W. G. & Whittet, D. C. B.) 592-597 (Astron. Soc. Pacif. Conf. Ser. 97, Bookcrafters, San Francisco, 1996).

7. Zaritsky, D. & Lo, K. Y. *Astrophys. J.* **303**, 66–75 (1986).
8. Ishizuki, S. *et al.* *Astrophys. J.* **355**, 436–441 (1990).
9. Regan, M. W. & Vogel, S. N. *Astrophys. J.* **452**, L21–L24 (1995).
10. Rogstad, D. H., Shostak, G. S. & Rots, A. H. *Astron. Astrophys.* **22**, 111–119 (1973).
11. Tacconi, L. J. & Young, J. S. *Astrophys. J.* **308**, 600–610 (1986).
12. Carignan, C., Charbonneau, P., Boulangier, F. & Viallefond, F. *Astron. Astrophys.* **234**, 43–52 (1990).
13. Binney, J. & Tremaine, S. *Galactic Dynamics* (Princeton Univ. Press, 1987).
14. Lin, C. C. & Shu, F. H. *Astrophys. J.* **140**, 646–655 (1964).
15. Lin, C. C. & Shu, F. H. *Proc. Natl Acad. Sci. USA* **55**, 229–234 (1966).
16. Athanassoula, E. *Phys. Rep.* **114**, 319–403 (1984).
17. Roberts, W. W. Jr & Yuan, C. *Astrophys. J.* **161**, 887–902 (1970).
18. Krause, M. in *The Cosmic Dynamo* (eds Krause, F., Rüdiger, K. H. & Rüdiger, G.) 305–310 (Int. Astron. Union Symp. No. 157, Kluwer Academic, Dordrecht, 1993).
19. Neinger, N., Beck, R., Sukumar, S. & Allen, R. J. *Astron. Astrophys.* **274**, 687–698 (1993).
20. Goldreich, P. & Lynden-Bell, D. *Mon. Not. R. Astron. Soc.* **130**, 125–158 (1965).
21. Bertin, G. & Lin, C. C. *Spiral Structure in Galaxies: A Density Wave Theory* (MIT Press, 1996).
22. Bardeen, J. M. in *Dynamics of Stellar Systems* (ed. Hayli, A.) 297–320 (Int. Astron. Union, Symp. No. 69, Reidel, Dordrecht, 1975).
23. Shu, F. H. *Astrophys. J.* **160**, 99–112 (1970).
24. Toomre, A. *Annu. Rev. Astron. Astrophys.* **15**, 437–478 (1977).
25. Chandrasekhar, S. & Fermi, E. *Astrophys. J.* **118**, 116–141 (1953).
26. Lynden-Bell, D. *Observatory* **86**, 57–60 (1966).
27. Lou, Y. Q. *Mon. Not. R. Astron. Soc.* **279**, L67–L70 (1996).
28. Miller, R. H., Prendergast, K. H. & Quirk, W. J. *Astrophys. J.* **161**, 903–916 (1970).
29. Ostriker, J. P. & Peebles, P. J. E. *Astrophys. J.* **186**, 467–480 (1973).
30. Vandervoort, P. *Astrophys. J.* **162**, 453–462 (1970).
31. Parker, E. N. *Astrophys. J.* **401**, 137–145 (1992).
32. Chakrabarti, S. K., Vainstein, S. I. & Rosner, R. *Nature* **368**, 434–436 (1994).
33. Mestel, L. & Subramanian, K. *Mon. Not. R. Astron. Soc.* **248**, 677–687 (1991).
34. Otmianowska-Mazur, K. & Chiba, M. *Astron. Astrophys.* **301**, 41–54 (1995).

ACKNOWLEDGEMENTS. This work was supported by the US NSF and NASA.

CORRESPONDENCE should be addressed to Y.-Q.L. (e-mail: lou@oddjob.uchicago.edu).

Fluorescence intermittency in single cadmium selenide nanocrystals

M. Nirmal*[†], B. O. Dabbousi*, M. G. Bawendi*, J. J. Macklin[†], J. K. Trautman[†], T. D. Harris[†], & L. E. Brus[†]

* Massachusetts Institute of Technology, Dept of Chemistry, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

[†] Bell Laboratories, Lucent Technologies, 600 Mountain Avenue, Murray Hill, New Jersey 07974, USA

SEMICONDUCTOR nanocrystals offer the opportunity to study the evolution of bulk materials properties as the size of a system increases from the molecular scale^{1,2}. In addition, their strongly size-dependent optical properties render them attractive candidates as tunable light absorbers and emitters in optoelectronic devices such as light-emitting diodes^{3,4} and quantum-dot lasers^{5,6}, and as optical probes of biological systems⁷. Here we show that light emission from single fluorescing nanocrystals of cadmium selenide under continuous excitation turns on and off intermittently with a characteristic timescale of about 0.5 seconds. This intermittency is not apparent from ensemble measurements on many nanocrystals. The dependence on excitation intensity and the change in on/off times when a passivating, high-bandgap shell of zinc sulphide encapsulates the nanocrystal^{8,9} suggests that the abrupt turning off of luminescence is caused by photo-ionization of the nanocrystal. Thus spectroscopic measurements on single nanocrystals can reveal hitherto unknown aspects of their photophysics.

In Fig. 1a we show a fluorescence image of single 21-Å-radius CdSe nanocrystals, embedded in a thin polyvinylbutyral film, at room temperature. The streaks in the image are a consequence of raster scanning the sample across a diffraction-limited laser spot, and arise due to the discrete turning on and off of the nanocrystal

fluorescence signal while acquiring the image (~2 min). A fluorescence intensity versus time trace for a single nanocrystal under continuous wave excitation is shown in Fig. 1b.

To determine whether this on/off phenomenon is light-induced or occurs spontaneously we investigated the intensity dependence of the on/off times. If the ons/off were due to spontaneous transformations between an emitting and a 'dark' configuration, the on/off periods would be independent of excitation intensity¹⁰. On the other hand, for a light-induced process where the 'dark' state is fed via a radiationless transition from the optically excited state, the 'on' times should decrease with increasing excitation intensity whereas the 'off' times would be determined by the spontaneous lifetime of the 'dark' state¹¹. Fluorescence intensity versus time traces at two excitation intensities (Fig. 2a) indicate that the average on time is inversely proportional to the excitation intensity while the average off time is intensity independent, suggesting a light-induced process.

This on/off behaviour is consistent with a mechanism based on Auger ionization. In nanocrystals with two or more electron-hole pairs, the energy released from the annihilation of a pair may be transferred to the remaining carriers, one of which is preferentially ejected into the surrounding matrix depending on the conduction (valence) band offset at the nanocrystal interface and the electron (hole) confinement energy within the nanocrystal. In the resulting ionized nanocrystal, the Auger interaction, which is mediated by the Coulomb repulsive potential between carriers, is strongly enhanced owing to the diminished screening relative to neutral nanocrystals. Subsequent photogenerated electron-hole pairs thus recombine primarily non-radiatively, transferring their energy to the resident carrier. This mechanism accounts for the decrease in the quantum efficiency of similar CdS nanocrystals embedded in glass on laser irradiation (photodarkening)¹². Photodarkened particles exhibit significantly shorter

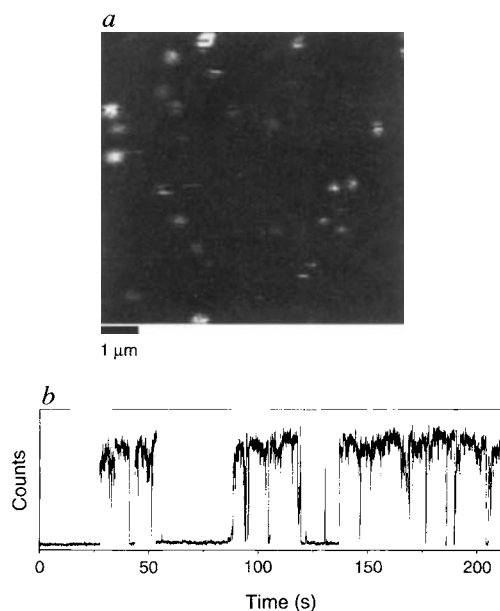


FIG. 1a, Image of a random field of single ~21-Å-radius CdSe nanocrystals (each with ~4 monolayers of ZnS on its surface) acquired by raster scanning the sample across a diffraction-limited laser spot (wavelength 532 nm, full-width at half-maximum ~0.38 μm) and collecting the red-shifted fluorescence onto an avalanche photodiode in an epi-illumination confocal geometry¹⁹. This 160 × 160 pixel image (1 pixel = 5 ms) represents an 8 × 8 μm field of view. Sample preparation involves spin-coating one drop of a ~5 nM solution of nanocrystals in a 0.5% polyvinylbutyral/toluene mixture onto a quartz cover slip. b, Fluorescence intensity versus time trace of a single ~21-Å-radius CdSe nanocrystal, from the same sample used in a, recorded using a multichannel scaler with a 40-ms sampling interval and an excitation intensity of ~0.52 kW cm⁻².

[†] Present addresses: Department of Chemistry, Columbia University, New York, New York 10027, USA (M.N., L.E.B.); seQ Ltd, Princeton, New Jersey 08540-6449, USA (J.J.M., J.K.T., T.D.H.).

emission lifetimes and lower quantum yield relative to neutral nanocrystals¹³. The luminescence is then restored when the ejected carrier returns to neutralize the particle. At low flux (about three orders of magnitude lower than in our study) the Auger ionization rate was observed to scale as the square of the excitation intensity¹². We observe that the average on period, which is determined by the ionization rate, scales linearly with intensity (Fig. 2a). Furthermore, mode-locked pulsed laser excitation at the same average power but with over two orders of magnitude higher probability of generating two electron-hole pairs, yielded similar on times as were found with continuous-wave excitation. This suggests a two-step ionization mechanism. The first step involves trapping of one of the carriers onto the nanocrystal surface followed by photoionization within the trap state lifetime ($\sim 5 \mu\text{s}$). At low flux both processes should scale linearly with excitation intensity (I), the quantum yield of the ionization step being $[(I\sigma_p)/(h\nu)] / \{[(I\sigma_p)/(h\nu)] + \gamma_i\}$, where σ_p is the photoionization cross-section and γ_i is the relaxation rate from the intermediate trap state. Saturation of either step would lead to a net linear dependence. For instance, at an excitation rate of $\sim 1 \text{ MHz}$ as in our study, a trap lifetime of $\sim 5 \mu\text{s}$ and assuming an ionization yield of $\sim 10\text{--}100\%$, the second step would approach saturation resulting in an effective three-level system with near-linear intensity dependence.

Based on this model the on/off times, determined by the photoionization/neutralization rates respectively, should be sensitive to changes at the nanocrystal-matrix interface. The nanocrystals as initially synthesized ('bare') have organic capping groups which passivate $\sim 1/2$ of all the surface atoms^{14,15}. These particles can instead be overcoated with a shell of ZnS, a higher-bandgap material^{8,9}, with the organic capping groups on the ZnS surface. In Fig. 2b, the fluorescence intensity versus time trace of a 'bare' particle is compared at the same excitation intensity with that of an overcoated one (with ~ 7 monolayers of ZnS). ZnS has a bandgap $\sim 2 \text{ eV}$ higher than that of CdSe and so should serve as an effective barrier to ionization/neutralization and passivate potential surface trap sites. This is apparent in the dramatic increase in both the average on/off times for the ZnS overcoated nanocrystal relative to the 'bare' particle. To confirm this, we probed on/off times at the same excitation intensity as a function of shell thickness for CdSe nanocrystals with 0, ~ 0.25 , ~ 2 and ~ 7 monolayers of ZnS. Strongly non-exponential histograms of on/off times for the four samples are shown in Fig. 3a and b. Each histogram represents on/off statistics for 18 nanocrystals. The

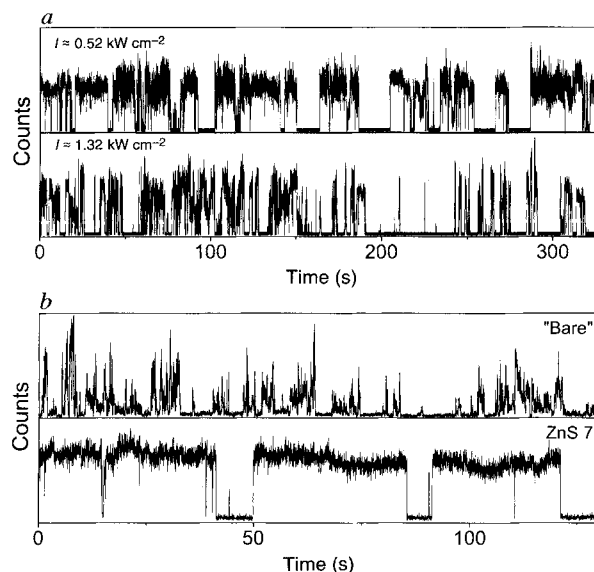


FIG. 2 a, Comparison of fluorescence-intensity versus time traces at excitation intensities (I) of ~ 0.52 and $\sim 1.32 \text{ kW cm}^{-2}$ with a sampling interval of 10 ms. The average on/off times at the two excitation intensities are: ($I \approx 0.52 \text{ kW cm}^{-2}$; $\langle\tau_{\text{on}}\rangle \approx 0.97 \text{ s}$, $\langle\tau_{\text{off}}\rangle \approx 0.44 \text{ s}$; $I \approx 1.32 \text{ kW cm}^{-2}$; $\langle\tau_{\text{on}}\rangle \approx 0.32 \text{ s}$, $\langle\tau_{\text{off}}\rangle \approx 0.43 \text{ s}$). Intensity-dependent studies were carried out on CdSe nanocrystals with ~ 4 monolayers of ZnS on their surface. b, Fluorescence-intensity versus time trace of a 'bare' nanocrystal compared with that of an overcoated one with a shell thickness of ~ 7 monolayers of ZnS (ZnS 7) at the same excitation intensity ($I \approx 0.70 \text{ kW cm}^{-2}$) and a sampling interval of 20 ms.

average off time yields the average neutralization time ($\langle\tau_n\rangle$) and the photoionization branching ratio, σ_i , defined as the ionization probability per excitation, is determined from the product of the excitation rate and the average on time. Consistent with photoionization/neutralization across a barrier, $\langle\tau_n\rangle$ increases and σ_i decreases with increasing shell thickness. A simple resonant tunnelling mechanism through the ZnS shell would exhibit a similar trend. Based on known band offsets between CdSe and ZnS and the electron/hole confinement energies, simple resonant tunnelling however predicts a relative branching ratio between the 'bare' and the 7-monolayer ZnS-shell sample that is two orders of

FIG. 3 a, Histograms of on times for CdSe nanocrystals overcoated with 0, 0.25, 2 and 7 monolayers of ZnS. The ZnS coverage was determined using quantitative X-ray spectroscopy (WDS) and small-angle X-ray scattering (SAXS). Individual histograms include statistics from 18 nanocrystals with each nanocrystal being probed for 2.73 min with a 20-ms sampling interval and an excitation intensity (I) of $\sim 0.70 \text{ kW cm}^{-2}$. The average on times ($\langle\tau_{\text{on}}\rangle$) for the four samples in order of increasing shell thickness are 0.2, 0.28, 0.62 and 1.52 s with corresponding photoionization branching ratios ($\sigma_i = h\nu/\sigma\langle\tau_{\text{on}}\rangle$) of 5.1×10^{-6} , 3.6×10^{-6} , 1.6×10^{-6} and 6.7×10^{-7} respectively. The absorption cross-section, σ , for these nanocrystals ($\sim 5.2 \times 10^{-16} \text{ cm}^2$ at a wavelength of 532 nm) was obtained from solution measurements. b, Histograms of off times for the four samples from the same dataset as in a. The average off (neutralization) times ($\langle\tau_n\rangle$) in order of increasing shell thickness are 0.29, 0.31, 0.74 and 1.24 s.

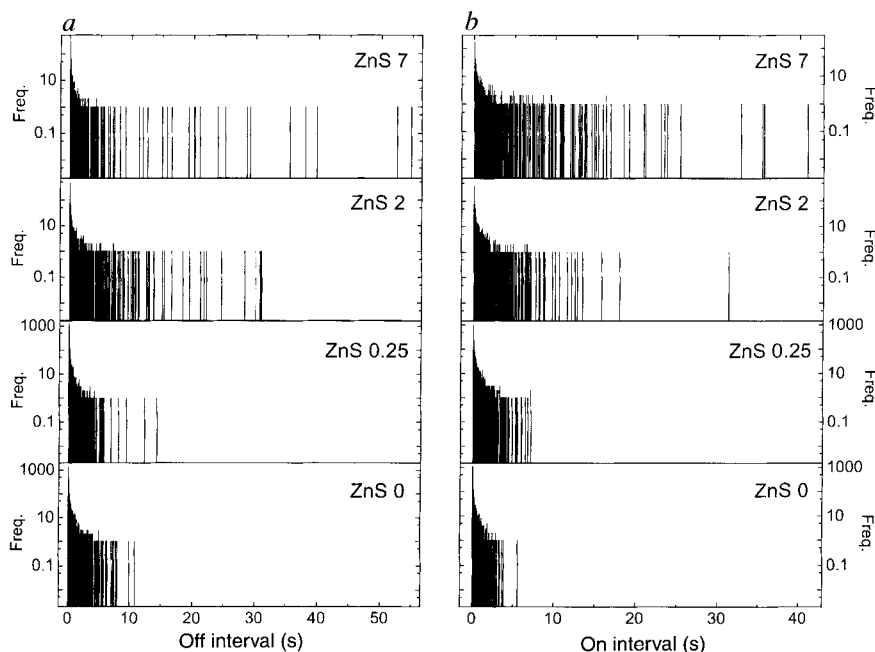
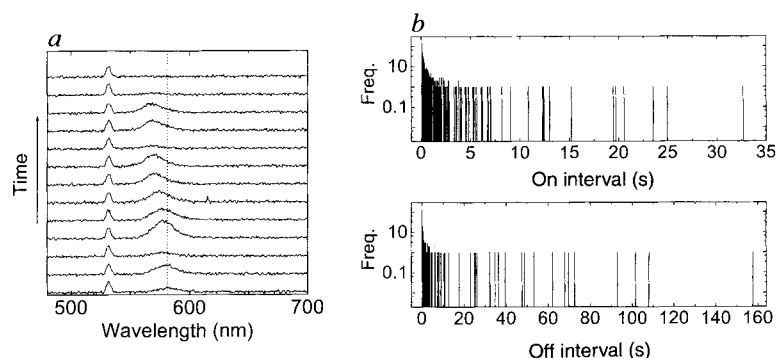


FIG. 4a, Fluorescence spectra (1 min integration per spectrum) of a single nanocrystal as a function of time. The sharper feature to the blue (wavelength $\lambda = 532$ nm) is residual laser light which is included for reference. Fluctuations in fluorescence intensity relative to the laser are due to the discrete turning on and off of the emission during the experiment. Assuming an effective mass model for the quantum-confined energy levels of the nanocrystal²⁰, the observed blue-shift with time (~ 50 meV) corresponds to a ~ 1.75 -Å shrinkage in the radius representing $\sim 2/3$ of a monolayer of CdSe. The fluorescence signal vanishes permanently in the thirteenth minute. b, Histograms of on/off times for a single 'robust' nanocrystal (see text) probed for ~ 39 min with a sampling interval of 40 ms and an excitation intensity of ~ 0.53 kW cm⁻².



magnitude larger than that observed experimentally. An 'over-the-barrier' process such as Auger ionization should be significantly less sensitive to shell thickness.

Although the nanocrystals are ionized and then neutralized reversibly, single-nanocrystal spectra shift irreversibly to the blue (Fig. 4a). As the quantum-confined energy levels vary as $1/\text{radius}^2$, this blue-shift indicates a shrinkage of the effective radius as a function of time and is most probably due to surface photooxidation¹⁶. This probably causes nanocrystals to photobleach permanently before we can accumulate accurate single-dot statistics. But sometimes we find 'robust' nanocrystals which we have been able to study for up to 40 minutes under constant illumination at room temperature. In Fig. 4b we show the on/off histograms of one such nanocrystal. Both these distributions, which are also highly non-exponential, strongly suggest that even a single nanocrystal exhibits distributed kinetics, in sharp contrast to a single molecule which usually has a single exponential intersystem crossing rate and triplet lifetime^{17,18}. A 21-Å-radius nanocrystal consists of $\sim 1,400$ atoms, of which $\sim 25\%$ lie on the surface. Surface reconstructions, defects and progressive oxidation could potentially lead to a range of ionization/neutralization rates.

The on/off kinetics, obscured in previous ensemble studies, reveal that although ionization is rare (probability per excitation is $\sim 10^{-6}$), it has a significant effect on luminescence count rates owing to the unusually long ($\tau_n \approx 0.5$ s) neutralization times that follow. This poses a potential limitation to optical devices based on CdSe nanocrystals for which fast response times and high luminescence efficiency are crucial. Approaches for modifying

this on/off behaviour could include the use of an inorganic shell material with lattice matching superior to ZnS. Alternatively, because the on/off behaviour currently observed depends on the degree of stabilization of the ejected carrier within the matrix, this phenomenon could conceivably be exploited as a sensitive probe (especially in biological systems) of the polarity of the local environment.

Received 1 July; accepted 24 September 1996.

1. Alivisatos, A. P. *Science* **271**, 933–937 (1996).
2. Brus, L. E. *Appl. Phys. A* **53**, 465–474 (1991).
3. Colvin, V. L., Schlamp, M. C. & Alivisatos, A. P. *Nature* **370**, 354–357 (1994).
4. Dabbousi, B. O., Bawendi, M. G., Onitsuka, O. & Rubner, M. F. *Appl. Phys. Lett.* **66**, 1316–1318 (1995).
5. Weisbuch, C. J. *Cryst. Growth* **138**, 776–785 (1994).
6. Kirstaedter, N. et al. *Appl. Phys. Lett.* (in the press).
7. Schröck, E. et al. *Science* **273**, 494–497 (1996).
8. Hines, M. A. & Guyot-Sionnest, P. *J. Phys. Chem.* **100**, 468–471 (1996).
9. Kortan, A. R. et al. *J. Am. Chem. Soc.* **112**, 1327–1332 (1990).
10. Ambrose, W. P., Basche, T. & Moerner, W. E. *J. Chem. Phys.* **95**, 7150–7163 (1991).
11. Bernard, J., Fleury, L., Talon, H. & Orrit, M. *J. Chem. Phys.* **98**, 850–859 (1993).
12. Chepic, D. I. et al. *J. Lumin.* **47**, 113–127 (1990).
13. Roussignol, P., Ricard, D., Lukasik, J. & Flytzanis, C. *J. Opt. Soc. Am. B* **4**, 5–13 (1987).
14. Murray, C. B., Norris, D. J. & Bawendi, M. G. *J. Am. Chem. Soc.* **115**, 8706–8715 (1993).
15. Becerra, L. R., Murray, C. B., Griffin, R. G. & Bawendi, M. G. *J. Chem. Phys.* **100**, 3297–3300 (1994).
16. Bowen Katari, J. E., Colvin, V. L. & Alivisatos, A. P. *J. Phys. Chem.* **98**, 4109–4117 (1994).
17. Basche, T., Kummer, S. & Brauchle, C. *Nature* **373**, 132–134 (1995).
18. Cook, R. J. & Kimble, H. J. *Phys. Rev. Lett.* **54**, 1023–1026 (1985).
19. Macklin, J. J., Trautman, J. K., Harris, T. D. & Brus, L. E. *Science* **272**, 255–258 (1996).
20. Norris, D. J. & Bawendi, M. G. *Phys. Rev. B* **53**, 16338–16346 (1996).

ACKNOWLEDGEMENTS. We thank Al. L. Efros for extensive conversations.

CORRESPONDENCE should be addressed to M.N. (e-mail: nimal@chem.columbia.edu).

Efficient real-time confocal microscopy with white light sources

R. Juškaitis, T. Wilson, M. A. A. Neil & M. Kozubek

Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK

THE main advantage of confocal microscopes over their conventional counterparts arises from their ability to optically 'section' nearly transparent materials; the thin image slices thus obtained can be used to reconstruct three-dimensional images, a capability which is particularly useful for the study of biological specimens. Confocal microscopes have previously used either a single laser-illuminated point-source and single point-detector (which are scanned in tandem across the object) or white-light illumination with multiple point-sources and detectors. Single-point-source systems, however, do not usually form images in real time and are restricted to using available laser wavelengths. Multiple-point-

source systems, on the other hand, produce images in real time but use light very inefficiently—typically 1% or less is used for imaging. Here we demonstrate a white-light, multiple-point-source method which can in principle produce images in real time with light efficiencies as high as 50%. This system is likely to find broad practical application, particularly in the imaging of weakly reflecting or weakly fluorescent specimens.

In order to build a confocal microscope, it is necessary to ensure that the illumination and detection systems are arranged such that light which has originated from a specific position in the source plane is detected only at the equivalent position in the detector plane¹. This is trivial to achieve in single pinhole source/detector systems where scanning is also required to image a finite region of the object. In multiple point, tandem scanning reflection systems² the pinholes must be well spaced to minimize crosstalk between neighbouring 'confocal systems' and this leads to inefficient use of the available light—typically only 0.5–1.0% is used. It is also necessary to scan in order to fill in the gaps between portions of the object probed by the well spaced 'confocal systems'. Our approach is fundamentally different and does not require scanning. Let us consider the reflection system of Fig. 1. The key element is the aperture mask which consists of an array of apertures which are packed as closely together as possible in order

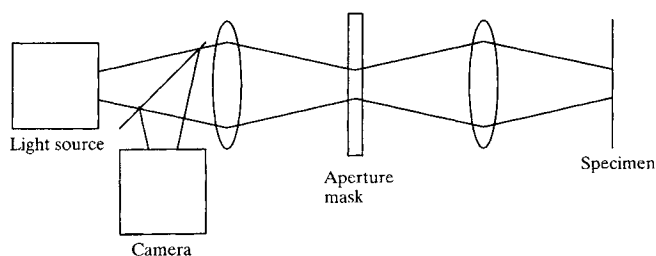


FIG. 1 Schematic diagram of the optical system.

to use all of the available light. The aperture mask is also programmable, in the sense that the transmissivity of individual apertures may be modulated according to a variety of time dependent sequences. We prevent crosstalk by ensuring that the sequences are chosen such that the pattern of openings and closings of different apertures are completely uncorrelated with one another. A reasonable approximation to this would be to use independent binary random sequences of 1 and -1 for each aperture transmissivity or, perhaps, time shifted complementary Golay sequences as these are known to possess zero cross correlation for finite sequence lengths³. The practical difficulty here is that the codes we have mentioned require negative numbers which cannot be achieved optically and so it is necessary to introduce a constant baseline shift. This means that the final image will consist of a confocal image superimposed on a conventional image. It is, however, a reasonably simple matter to subtract out the conventional image electronically, in real time. This is the key to our approach. We do not try to obtain a pure confocal image directly but rather obtain a composite image which permits us to use masks with greater light efficiency than those used in the tandem scanning system.

To achieve real-time imaging, it is necessary that the transmissive states of the apertures in the mask change quickly. One method is to use a fast spatial light modulator consisting of an array of pixels each of which is individually addressable. Alternatively, it is possible to build an aperture mask onto which the sequences or modulation codes are impressed photolithographically. If the mask is now rotated, the pattern presented in time to any fixed position in the detector plane is equivalent to that impressed in space along the corresponding arc of the disk. In both cases a composite confocal and conventional image is produced.

We elected to use a rotating disk in our system. Square apertures ($80 \times 80 \mu\text{m}$) on an $80\text{-}\mu\text{m}$ pitch were photolithographically defined in aluminium on a perspex disk. The aperture transmission was either zero or unity and the pattern of transmissive apertures was randomly distributed over a sector of the disk. A complementary pattern was impressed on the adjacent sector such that the track along any given arc sees an average transmission of exactly 0.5. Because it is necessary to subtract a conventional image from the composite image to provide the pure confocal image, we also provided blank a sector on the disk to produce the conventional image. The disk was rotated at approxi-

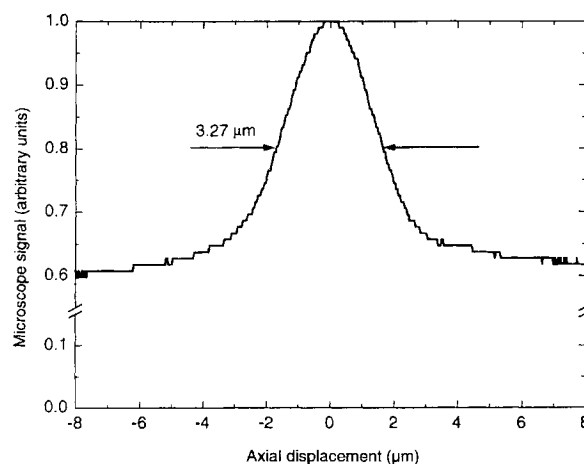


FIG. 2 The axial response obtained by measuring the image signal detected through an encoded sector of the disk as a plane mirror is scanned through focus. A $\times 32$, 0.5 NA Leitz objective was used together with 670-nm-wavelength light from a 250 W tungsten-halogen projector lamp. The response consists of the superposition of a confocal response, which exhibits optical sectioning, together with a constant conventional response which does not exhibit optical sectioning. The mirror was scanned piezo-electrically and the axial displacement was calibrated interferometrically.

mately 1,500 revolutions per minute (r.p.m.) and synchronized with the frame rate of the camera used.

To demonstrate the composite behaviour of our system, we scanned a perfect reflector, axially through focus, to measure the signal detected through an encoded sector of the disk. Because a conventional microscope does not exhibit optical sectioning, we would expect that the combined response would consist of a confocal axial response superimposed on a constant baseline due to the conventional contribution. Figure 2 shows the response we obtained using a $\times 32$, 0.5 numerical aperture (NA) Leitz objective. The light source was a 250-W tungsten-halogen incandescent projector lamp together with the a narrow-band red filter ($\lambda = 670 \text{ nm}$, $\Delta\lambda = 10 \text{ nm}$). The confocal axial response is clearly seen and has a full-width at half-maximum of $3.27 \mu\text{m}$ which compares well with a theoretical prediction of $3.19 \mu\text{m}$ obtained by modelling the system using the theory of Wilson and Hewlett⁴, originally developed for tandem scanning microscopes.

The axial response (Fig. 2) confirms that a pure confocal response may be extracted from the composite response by subtracting the constant signal due to the conventional image. To demonstrate that a pure confocal image could be extracted from the composite image, we decided to concentrate on the three-dimensional structure of osteoclast resorption pits in sperm whale dentine, as this is a weakly reflecting specimen. A CCD (charge-coupled device) camera was used to capture both the composite and conventional images. Figure 3 shows the three-dimensional structure of the resorption pit illustrating the previous activity of an osteoclast on the mineralized tissue. We stress

FIG. 3 Complex resorption pit excavated by an osteoclast in a slice of sperm whale dentine. The image was obtained with a $\times 50$, 0.8 NA Olympus objective together with a narrowband green filter ($\lambda = 546 \text{ nm}$; $\Delta\lambda = 10 \text{ nm}$). The field size is $150 \times 150 \mu\text{m}$. The depth of the deepest pit is measured to be $10 \mu\text{m}$. The image was obtained by combining height information with an auto-focus image obtained from 60 through-focus confocal image slices spanning $15 \mu\text{m}$. In this image, colour represents depth and the brightness of the colour is determined by the intensity of the autofocus image. Our image capture system consisted of a PC together with a DT3155 Data Translation imaging board which allowed us to perform near real-time subtraction of the conventional image from the composite image and to image the confocal slices at the rate of 17 frames per second.

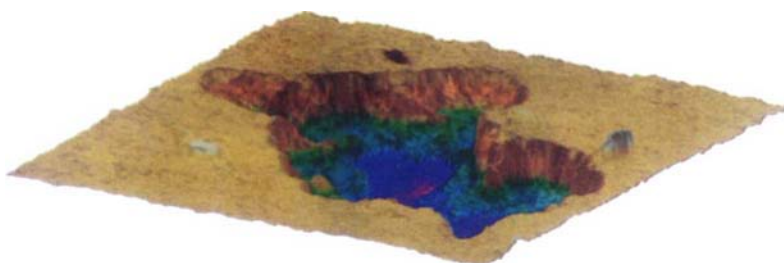
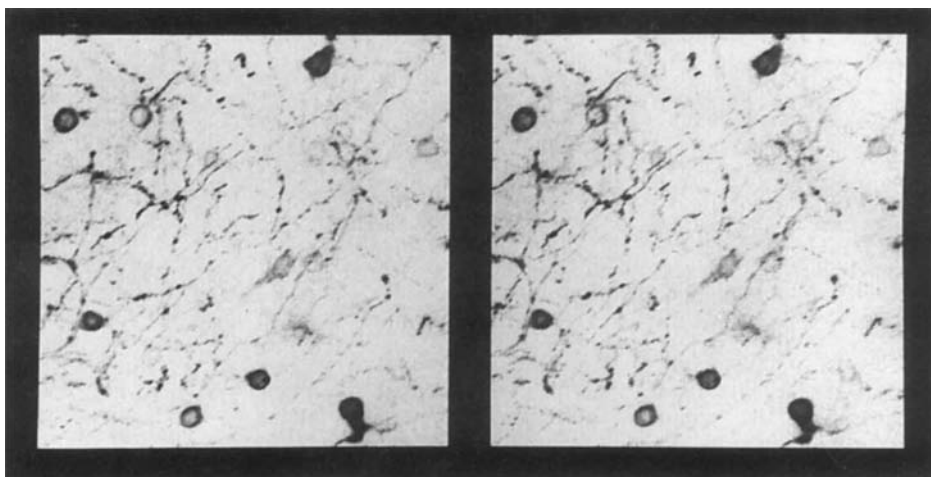


FIG. 4 Autofocus stereo pairs obtained from 18 through-focus sections obtained at 5- μm intervals from the bulk of a 100- μm -thick coronal section in the thalamus region of a two-day-old mouse brain. The primary somatosensory cortex had been labelled with DiI. The dye was then passively transported along the axons to the thalamus during a 4-week incubation period at 37 °C. A Leica $\times 100$ 1.4 NA objective lens was used, together with a 488-nm interference filter for the excitation illumination and a 560-nm long-pass filter for the fluorescence signal. The only other modification to the optical system in this case was the use of a cooled CCD camera. The field size is 100 $\times$ 100 μm .



that these images were taken with an inexpensive CCTV camera and that, even so, the illumination required attenuation owing to the highly favourable light budget. In other work⁵, using a tandem scanning microscope to image such samples, it proved necessary to use an arc lamp and to prepare the sample as for SEM imaging. No such treatment or light source was required to obtain this image.

In our system the encoded sectors contained an equal number of transmissive and non-transmissive aperture and hence the confocal image is formed using 25% of the available light. This should be compared with the 0.5–1% light budget of the tandem scanning instruments. Using ferroelectric liquid-crystal spatial light modulators and polarization effects to separate the confocal from the combined image, a 50% light budget can be obtained.

To show the capability of our system to image weakly fluorescent specimens we imaged a 100- μm -thick coronal section of mouse brain. A single crystal of carbocyanine dye, DiI (1,1'-didioctadecyl-3,3',3'-tetramethylindocarbocyanine percholate; Molecular Probes, Inc., Eugene, Oregon, USA) was placed on the primary somatosensory cortex of a fixed, two-day post-natal mouse brain. The dye was passively transported back along axons to their originating cell bodies located in the thalamus region of the brain. The fluorescence was excited by placing a 488-nm interference filter (10-nm bandwidth) in the illumination path and the image signal was measured by detecting wavelengths beyond 560 nm. A Leica 1.4 NA oil-immersion objective was used and the CCTV camera was replaced by a Princeton Instruments cooled CCD camera but the illumination was still provided by the slide projector bulb. Figure 4 shows an auto-focus stereo pair generated from images taken from a coronal section in the thalamus region.

We have shown that by relaxing the requirement to obtain a pure confocal image directly that a very simple, light-efficient microscope may be built. The key to our approach has been to implement aperture encoding techniques to ensure zero cross correlation between different apertures in the aperture mask. This removes the requirement that the apertures be placed far apart and hence provides for a highly light-efficient confocal microscope. Our approach also obviates the need to use laser illumination and may be incorporated very easily into existing conventional microscopes. □

Received 30 April; accepted 4 September 1996.

1. Wilson, T. (ed.) *Confocal Microscopy* (Academic, London, 1990).
2. Petráň, M., Hadravský, M., Egger, M. D. & Galambas, R. J. *Opt. Soc. Am.* **58**, 661–664 (1968).
3. Golay, M. J. E. *J. Opt. Soc. Am.* **39**, 437–444 (1949).
4. Wilson, T. & Hewlett, S. J. *J. Microsc.* **163**, 131–150 (1991).
5. Boyde, A. in *Electronic Light Microscopy* (ed. Shotton, D.) (Wiley-Liss, New York, 1993).

ACKNOWLEDGEMENTS. We thank A. Boyde for the sperm whale dentine specimen and R. Adams and Z. Molnar for the mouse brain. M.K. was supported by the Soros Foundation. This work is supported by the Paul Instrument Fund.

CORRESPONDENCE should be addressed to T.W. (e-mail: tony.wilson@eng.ox.ac.uk).

Rapid changes in ocean circulation and heat flux in the Nordic seas during the last interglacial period

Torben Fronval* & Eystein Jansen

Department of Geology, University of Bergen, Allegaten 41, 5007 Bergen, Norway

THE apparent similarity of climate variability in the North Atlantic region in the last interglacial period^{1–5} and the present interglacial (Holocene) has recently been challenged by the rapid oscillations in climate conditions indicated by some marine^{6,7} and terrestrial climate records^{8–10} for the last interglacial. Ocean circulation in the northern North Atlantic seems to be intimately coupled to the processes of climate change on various time-scales^{11,12}, so that climate variability—and the associated mechanisms of change—should be well recorded by sediments from these high latitudes. Previous studies in this region^{5,13} have indeed indicated an apparently less stable last-interglacial climate than at middle latitudes of the North Atlantic Ocean⁴. Here we present detailed records from marine sediments in the Nordic seas of oceanographic conditions during the last interglacial. The records show three large sea surface temperature fluctuations, a weakening of the east–west sea surface temperature gradient with time, and changes in deep-water properties. In contrast, similar analyses of a core from the same region indicate that sea surface temperature during the Holocene has been relatively stable. Our data—along with those from the Labrador Sea⁷—indicate rapid changes in ocean circulation and oceanic heat fluxes at high northern latitudes during the last interglacial, which may have been associated with marked temperature changes on adjacent continents.

The Nordic seas (the Norwegian, Iceland and Greenland seas) are situated within the northernmost limb of the North Atlantic thermohaline circulation and are characterized by strong gradients between warm Atlantic water masses and cold polar waters. Consequently, climate instability due to longitudinal and latitudinal displacement or changes in the strength of this circulation is likely to be reflected by the properties of the underlying marine

* Present address: Geological Institute, University of Copenhagen, Østervoldgade 10, 1350 Copenhagen K, Denmark.

sediment. Former reconstructions of the last interglacial period in the Nordic seas have indicated that warm conditions appeared as a single pulse of short duration^{5,13}, not as a long period of stable, warm climate conditions as found further south in the North Atlantic⁶.

We analysed three sediment cores, ODP site 644 (66° 40' N, 04° 34' N, 1,227 m water depth) from the Vöring plateau, HM71-19 (69° 29' N, 09° 31' N, 2,210 m) from the northern Iceland plateau and HM57-7 (68° 26' N, 13° 52' N, 1,620 m) from the western Iceland plateau (Fig. 1a). As the Holocene section of site 644 is incomplete, we used measurements from the Norwegian Sea core HM52-43 (63° 31' N, 00° 44' N, 2,781 m) to document the Holocene climate evolution. Former studies of these cores^{12,14-16} (see also Fig. 1b) have shown that global as well as regional climate variations are reflected in the sediment parameters.

In general, ages were assigned from the SPECMAP stack¹⁷ to isotopic events and stage boundaries identified in the planktonic oxygen-isotope records. In specifying the 6/5e and 5e/5d boundaries, the records of benthic oxygen isotopes, *Neogloboquadrina pachyderma* (s.) percentage, ice rafted detritus (IRD) content and foraminiferal content were taken into consideration (see Fig. 2 legend). The light values of $\delta^{18}\text{O}$ below 13.2 m depth in the core from site 644, below 142 cm in core HM71-19 and below 235 cm in core HM57-7 (Fig. 1b), all occurring during intervals of high IRD deposition, are attributed to influence of isotopically light deglacial water and are consequently interpreted as representing late stage 6 (termination II). Similarly light isotope signals are also common in the Nordic seas during termination I^{14,18}. Core depth was converted to age by linear interpolation between dated levels. The marine isotope substage 5e correlates to the Eemian on land¹⁹.

Silicic ash layers of substage 5e age have previously been described in cores from the Nordic seas^{20,21}. A silicic ash horizon was identified in the substage 5e section of all the investigated cores (Fig. 1b). The chemical composition of the ash grains are

very similar, which indicate that the tephra horizons represent the same event and can be used as a stratigraphic time marker. The occurrence of the tephra horizon around 122–121 kyr ago in all three cores thus supports the age models.

To construct high-resolution proxy temperature records we calculated the percentage of *N. pachyderma* (s.) in the >150- μm fraction (the 125- μm fraction in core HM57-7). The $\delta^{18}\text{O}$ composition of *N. pachyderma* (s.) tests also reflect changes in sea surface temperature (SST), at least in the temperature range of 2–8 °C (ref. 22), but the temperature-related shifts are partly masked by ice-volume and salinity changes, especially around terminations. During the main part of substage 5e, changes in *N. pachyderma* (s.) percentages and planktonic $\delta^{18}\text{O}$ are often parallel and we consider both as reflecting SST variations in this interval. The tandem behaviour provides a double check on the credibility of the SST variations.

Normally, as observed on the Iceland plateau (Fig. 2b and c), interglacial sediments from the high-latitude North Atlantic region are rich in foraminifers, reflecting the higher carbonate productivity during warm periods than during cold periods. At the Vöring plateau, however, proxy temperature and planktonic foraminiferal content data anticorrelate (Fig. 2a). Although dissolution of foraminifera tests is observed, it is unlikely that such dissolution is responsible for this anticorrelation, as the dominant 'warm' species, *Globigerina bulloides* and *Globigerina quinqueloba*, are regarded as having low resistance to solution compared to *N. pachyderma* (s.)²³. The low foraminiferal content in the warm intervals might thus be due to a combination of low carbonate productivity in the warm Atlantic waters compared to the Arctic surface-water masses, and masking by high input of fines after the previous glacial period.

The absolute content of IRD was counted to estimate the degree of iceberg discharge. In both the Norwegian and Iceland seas, ice-rafting associated with melting of the late Saalian ice sheet culminated in late stage 6 (134–130 kyr). Only limited ice-rafting is documented within the main part of substage 5e,

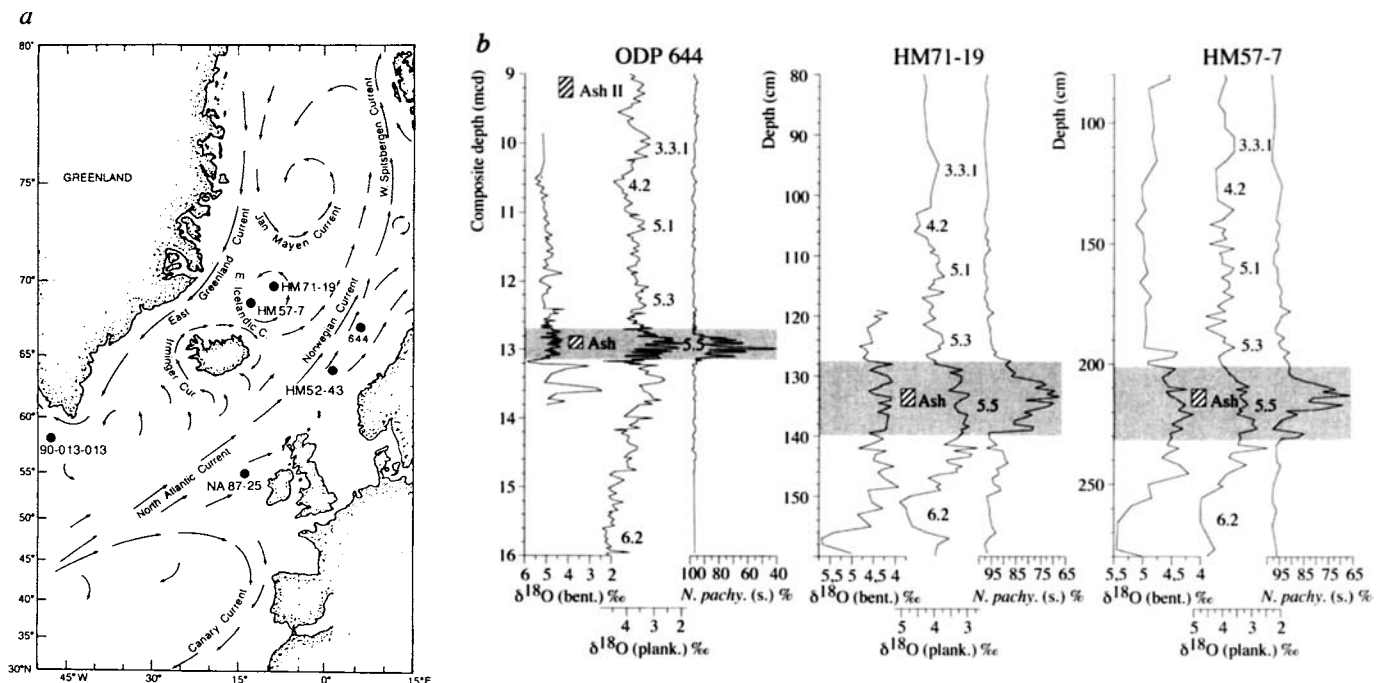


FIG. 1 a, Location of sediment cores and their position with respect to present surface water currents. b, Records of benthic and planktonic $\delta^{18}\text{O}$, and *N. pachyderma* (s.) percentage, from cores ODP 644, HM71-19 and HM57-7 covering the period 50–135 kyr. The composite depth scale for site 644 was formed by splicing together cores from holes 644A and 644B

on the basis of magnetic susceptibility records. The stratigraphic position of oxygen-isotope stages and substages (the last interglacial) is indicated by numbers. The position of substage 5e (the last interglacial) is marked by horizontal shaded bars. The level of Ash Zone II and the 'substage 5e' ash zone is shown by hatched boxes. The HM57-7 data are partly taken from ref. 15.

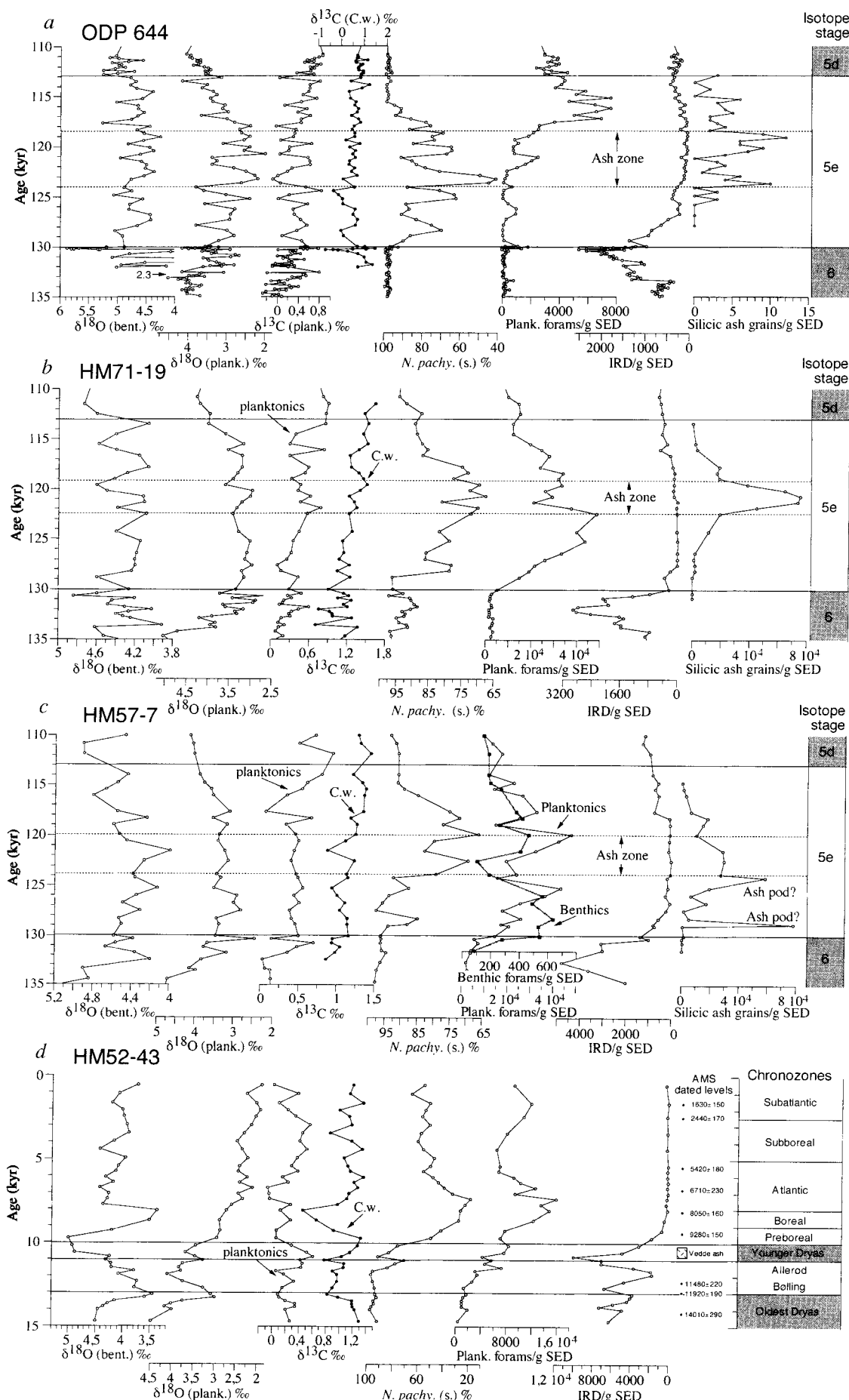
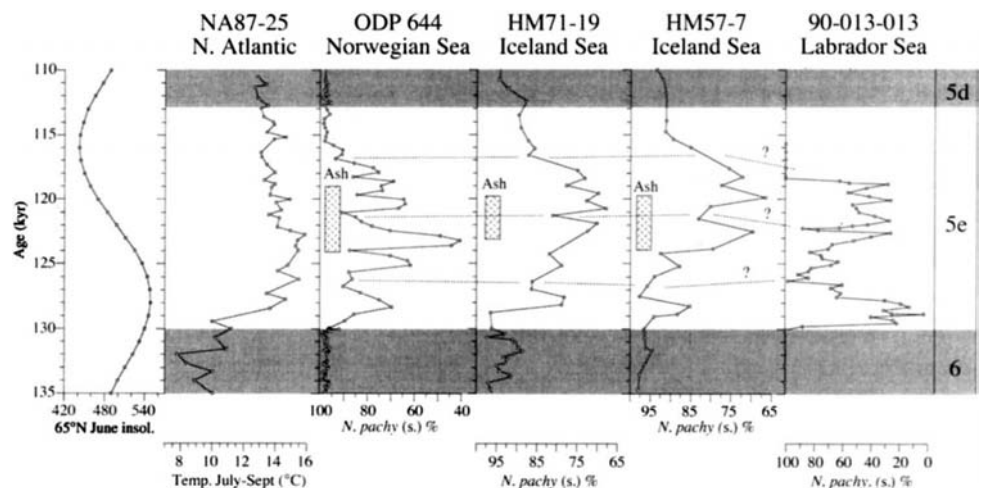


FIG. 2 a–c, Summary of climate proxies from cores ODP 644, HM71-19 and HM57-7 covering the period 135–110 kyr. Shown from left to right are benthic $\delta^{18}\text{O}$, planktonic $\delta^{18}\text{O}$, planktonic $\delta^{13}\text{C}$, $\delta^{13}\text{C}$ of *Cibicides wuellerstorfi* (C. w.), *N. pachyderma* (s.) percentage and foraminiferal content, IRD content and the content of silicic ash grains all expressed as number of grains per gram sediment (no./g SED). The stratigraphic position of isotope stages and substages are indicated by numbers. A multi-parameter approach is applied to assign the positions of the 6/5e and 5e/5d boundaries. The 6/5e boundary is placed between the last major glacial IRD peak and the first marked decrease in *N. pachyderma* (s.) percentage (SST increase). The 5e/5d transition is characterized by a gradual increase in $\delta^{18}\text{O}$ values (benthic and planktonic) and a gradual decrease in foraminiferal content. The 5e/5d boundary is placed at the centre of this transition. The location and vertical extension of the silicic ash zone, based on absolute ash-particle content and visual examination of the cores, is indicated by horizontal dashed lines. The two horizons of maximum ash-particle content in core HM57-7 either reflect the occurrence of isolated ash pods or represent a second and third ash zone of local extension. d, Summary of similar climate proxies from core HM52-43 covering the past 15 kyr. The age model is based on nine radiocarbon accelerator mass spectrometry (AMS) dates and the identification of the Vedde ash (age, 10,600 BP). The data are taken in part from ref. 14.

FIG. 3 Comparison of the insolation curve (65° N, June; extreme left)³⁴ and proxy temperature records from cores NA87-25⁵, ODP 644, HM71-19, HM57-7 and 90-013-013⁷. The levels of the silicic ash zone are marked by boxes filled with slanted lines. The age models of cores NA87-25 and 90-013-013 were created by correlation of the benthic and planktonic oxygen-isotope curves^{5,35}, respectively, to the SPECMAP timescale¹⁴. Dashed lines indicate possible correlation lines between the cores. The position of the cold (sub)stages 6 and 5d is indicated by horizontal shaded bars.



suggesting that no pronounced ice advances occurred anywhere around the Nordic seas. However, a small increase in IRD deposition occurs around 117 kyr, at the same time as the late 5e cooling of the entire basin.

The proxy temperature records from the Nordic seas display a number of high-frequency and high-amplitude variations (Fig. 2a–c). The number of temperature fluctuations and their shape are quite similar in the three records, suggesting that the temperature variations are in phase. Three main coolings (~127–126 kyr, ~122–121 kyr and ~117 kyr) and several minor coolings are recognized in the records. In comparison, only one main cooling (representing the termination of the Holocene climate optimum) is recognized in Holocene records from the Norwegian (Fig. 2d) and Iceland^{15,16} seas. This documents a different degree of stability in the two interglacials, at least in the Nordic seas region, with a relatively stable Holocene and an Eemian characterized by rapid temperature oscillations. During the warm intervals between 130 and 122 kyr, SSTs were much higher on the Vöring plateau than on the Iceland plateau, whereas the warm intervals after 121 kyr are characterized by more uniform SSTs basinwide. This weakening of the gradient could reflect an intensification of the Irminger current, a longitudinal displacement of the Norwegian current towards the west or a weakening of the East Greenland current.

Corresponding changes in SST and benthic $\delta^{18}\text{O}$, particularly at site 644, suggests coherent changes in SST and temperature/salinity of the intermediate waters. Thus, the data document a connection between rapid changes in deep- and surface-water properties, lending credence to the idea that the substage 5e instability was more than a local phenomenon.

Records of SST from the central North Atlantic indicate a relatively stable North Atlantic Drift^{4,5} during the last interglacial. However, close examination of core NA87-25 (Fig. 3) reveals a series of small SST fluctuations and one more pronounced cooling at 122–123 kyr, which may be coherent with the strong mid-Eemian cooling found at site 644. In core 90-013-013 from the southern Labrador Sea (Fig. 3), two major and several minor coolings are indicated by the *N. pachyderma* (s.) percentage record. The similarities in the number and calculated age of the SST fluctuations in the Nordic seas and the Labrador Sea indicate that at least the major coolings represent the same events. Weakened influence of Atlantic surface waters between 127 and 124 kyr and around 121 kyr is also observed in foraminiferal content data from the Greenland Sea⁶ and a marked mid-Eemian cooling is found in core V27-60 from the northern Norwegian Sea⁵. Together, these data indicate that coherent shifts in the strength of the Norwegian and the Irminger (the northern branch) currents, with the potential to influence climate

on a regional scale, occurred several times during the last interglacial period.

The occurrence of several, more-or-less distinct Eemian climate fluctuations over Europe is supported by lake sediment records from the French Massif Central¹⁰ and Hollerup, Denmark S. Björck *et al.*, unpublished data). Most European pollen records indicate limited Eemian climate variability, but a marked cooling event at ~125 kyr, which may correspond to the first pronounced cooling in our records, is recorded in pollen sequences from Grande Pile⁸ and Bispingen⁹ (lowered winter temperature). The degree of climate variability deduced from pollen data apparently varies for different transfer functions used^{3,8} and difficulties in identifying the oscillations may result from the insensitivity of most transfer functions to interglacial temperature fluctuations. Consequently, it is possible that the rapid changes in ocean heat flux had a significant effect on European climate. As the validity of the Greenland ice-core data from the last interglacial has been questioned²⁴, possible links to changes in air temperature over Greenland are not discussed here.

Difficulties in identifying Eemian temperature oscillations in the North Atlantic (and Europe) might also reflect damping of the oscillations at lower latitudes, as is the case with the termination of the Holocene climate optimum. This mid-Holocene transition is recorded in SST data from the Nordic seas (Fig. 2d and refs 16, 25) and the Denmark strait²⁶, and in pollen data from Greenland²⁷ and Northwest Europe²⁸, whereas temperature reconstructions from the central North Atlantic²⁹ and southern Europe²⁸ indicate rather uniform temperatures within the entire Holocene.

The apparent geographical distribution of the temperature changes indicates that the interglacial climate variability recorded in the Nordic seas is quite different from that of the glacial 'Dansgaard-Oeschger' cycles. It is also clear that the fluctuations are restricted to the northernmost parts of the North Atlantic thermohaline circulation but, nevertheless, have the potential to influence climate on the adjacent continents.

The marked shift in east–west SST gradient observed in the middle of substage 5e indicates that the direction and strength of different currents can vary substantially, probably reflecting wind-field variability within the interglacial period. Our results lend credence to the view that interglacials may exhibit high-amplitude climate instability at high northern latitudes, and that marked variations in the northward heat flux existed also in the absence of Northern Hemisphere ice-sheet forcing. Thus, causes for the temperature changes must be sought outside the possible influence of the continental ice sheets. A number of scenarios exist from simplified ocean atmosphere models^{30,31}, many of which focus on the potential effect of changes in the high-latitude hydrological cycle, but we are, however, far from understanding

the reasons for Eemian climate variability. Despite the difficulties in establish an accurate stratigraphy, our data favour the interpretation that the cooling terminating the last interglacial in the Nordic seas happened before significant ice-volume increase^{5,32}, and that the high latitudes of the Northern Hemisphere respond early to orbital forcing of the climate system, as predicted by the model of Imbrie *et al.*³³. □

Received 25 March; accepted 25 September 1996.

1. Woillard, G. M. *Quart. Res.* **9**, 1–21 (1978).
2. Menke, B. & Tynni, R. *Geol. Jb.* **A 76**, 1–117 (1984).
3. Guiot, J. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **80**, 49–69 (1990).
4. McManus, J. F. *et al.* *Nature* **371**, 326–329 (1994).
5. Cortijo, E. *et al.* *Nature* **372**, 446–449 (1994).
6. Sejrup, H. P., Hafliðason, H., Kristensen, D. K. & Johnsen, S. *J. Quat. Sci.* **10**, 385–390 (1995).
7. Seidenkrantz, M.-S., Kristensen, P. & Knudsen, K. L. *J. Quat. Sci.* **10**, 77–82 (1995).
8. Guiot, J. *et al.* *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **103**, 73–93 (1993).
9. Field, M., Huntley, B. & Müller, H. *Nature* **371**, 779–782 (1994).
10. Thouveny, N. *et al.* *Nature* **371**, 503–506 (1994).
11. Kellogg, T. B. *Boreas* **9**, 115–138 (1980).
12. Fronval, T., Jansen, E., Bloemendal, J. & Johnsen, S. *Nature* **374**, 443–446 (1995).
13. CLIMAP Project Members *Quat. Res.* **21**, 123–224 (1984).
14. Veum, T., Jansen, E., Arnold, M., Beyer, I. & Duplessy, J.-C. *Nature* **356**, 783–785 (1992).
15. Eide, L., Beyer, I. K. & Jansen, E. *J. Quat. Sci.* **11**, 115–124 (1996).
16. Flatøy, A. thesis, Univ. Bergen (1991).
17. Martinson, D. G. *et al.* *Quat. Res.* **27**, 1–29 (1987).

18. Jansen, E. & Erlenkeuser, H. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **49**, 189–206 (1985).
19. Turon, J. L. *Nature* **309**, 673–676 (1984).
20. Sejrup, H. P. *et al.* *J. Quat. Sci.* **4**, 109–114 (1989).
21. Kellogg, T. B. thesis, Columbia Univ. (1973).
22. Wu, G. & Hillaire-Marcel, C. *Geochim. Cosmochim. Acta* **58**, 1303–1312 (1994).
23. Berger, W. H. *Mar. Geol.* **8**, 111–138 (1970).
24. Peel, D. A. *Nature* **378**, 234–235 (1995).
25. Koc, N., Jansen, E. & Hafliðason, H. *Quat. Sci. Rev.* **12**, 115–140 (1993).
26. Kellogg, T. B. in *Climatic Changes on a Yearly to Millennial Basis* (eds Möner, N. A. & Karlén, W.) 123–133 (Reidel, Hingham, MA, 1984).
27. Funder, S. *Bulletin Grönlands Geologiske Undersøgelse* **129**, 1–66 (1978).
28. Huntley, B. & Prentice, I. C. in *Global Climates since the Last Glacial Maximum* (eds Wright, H. E. *et al.*) 136–168 (Univ. Minnesota Press, Minneapolis, 1993).
29. Ruddiman, W. F. & Mix, A. C. in *Global Climates since the Last Glacial Maximum* (eds Wright, H. E. *et al.*) 94–124 (Univ. Minnesota Press, Minneapolis, 1993).
30. Weaver, A. J. & Hughes, T. M. C. *Nature* **367**, 447–450 (1994).
31. Shaffer, G. & Bendtsen, J. *Nature* **367**, 354–357 (1994).
32. Duplessy, J. C. & Labeyrie, L. in *Start of a Glacial* (eds Kukla, G. J. & Went, E.) 173–183 (Springer, Berlin, 1992).
33. Imbrie, J. *et al.* *Paleoceanography* **7**, 701–738 (1992).
34. Berger, A. L. *Quat. Res.* **9**, 139–167 (1978).
35. Hillaire-Marcel, C., de Vernal, A., Bilodeau, G. & Wu, G. *Can. J. Earth Sci.* **31**, 63–89 (1994).

ACKNOWLEDGEMENTS. We thank S. Björck, H. P. Sejrup and H. Hafliðason for help and discussions; ODP for providing samples for this study; R. Sørås and O. Hansen for mass spectrometer operation; and O. Tumyr for supervision of the chemical analyses. This work was supported by the Environmental Program of the EC, The Norwegian Research Council, Bergen University and the Carlsberg Foundation.

CORRESPONDENCE should be addressed to T.F. (e-mail: torbernf@geo.geol.ku.dk).

A possible Late Cambrian vertebrate from Australia

Gavin C. Young*, Valya N. Karatajute-Talimaa† & Moya M. Smith‡

* Australian Geological Survey Organisation, PO Box 378, Canberra ACT 2601, Australia

† Lithuanian Institute of Geology, Sevcenkos 13, 2600, Vilnius, Lithuania

‡ Craniofacial Development, Division of Dentistry, UMDS, Guy's Hospital, London Bridge, London SE1 9RT, UK

THE fossil record of early vertebrates starts with certainty with the dermal armour of agnathan fish from the Early–Middle Ordovician of Australia^{1,2}. Recent controversial acceptance that conodonts³ and the fragments called *Anatolepis*^{4,5} may be vertebrate remains⁶, extends their fossil record back to the Late Cambrian. Now a new type of phosphatic skeleton from Australia shows a three-layered structure that indicates vertebrate affinity, but with several unique features not known in other vertebrates. The new evidence challenges the most widely accepted current theory for the development of the vertebrate skeleton^{7–11}, which assumes the odontode (skin denticle) to be the primitive patterning component. The Australian material provides an alternative model for early vertebrate dermal armour with which to assess the vertebrate-like hard tissues in conodonts^{12,13} and the dermal armour of *Anatolepis*^{4–6,14}.

The small broken plates (Fig. 1a) were acid-etched from a limestone with an abundant Late Cambrian conodont and trilobite fauna¹⁵ in the Georgina Basin, western Queensland. Superficially, they resemble the dermal ornament of the Ordovician pteraspidomorph agnathan *Porophoraspis* (Fig. 1b) from central Australia¹⁶, but are approximately coeval (Iverian–Franconian¹⁷) with the Late Cambrian occurrence of *Anatolepis*⁵, for which new histological evidence of vertebrate affinity includes dentine odontodes with a central pulp cavity⁶. The new Australian taxon has a completely different appearance and structure from *Anatolepis*. Dermal fragments are ornamented with equidimensional tubercles, the margins of which are interrupted by some of the evenly spaced pores opening through the external surface. A covering of tiny pores is also seen in Ordovician arandaspid agnathans from central Australia and Bolivia^{1,16}, suggesting that some form of pore-canal system may be primitive for vertebrates. However,

arandaspid tissue histology is radically different from that reported here, with tubercles superimposed on spongy bone¹. Fragments of the new material vary in thickness, with some showing three distinct tissue layers, the assumed complete condition. The granular texture of the middle layer contrasts with the compact tissue of the surface and basal layers where it is preserved (Fig. 2b, c).

The translucent superficial tissue is hypermineralized relative to the middle and basal layers, and is continuous over the whole external surface. It is finely laminated, with each lamina ending at the pore openings. These features imply sequential epithelial secretion of an enamel-like tissue. The evenly spaced pore openings are generally expanded to a funnel shape, which shows the clear layering of surrounding tissue (Fig. 2c). The tubercles may be rounded or flat (Figs 1a, 2a). The basal layer is rarely preserved,

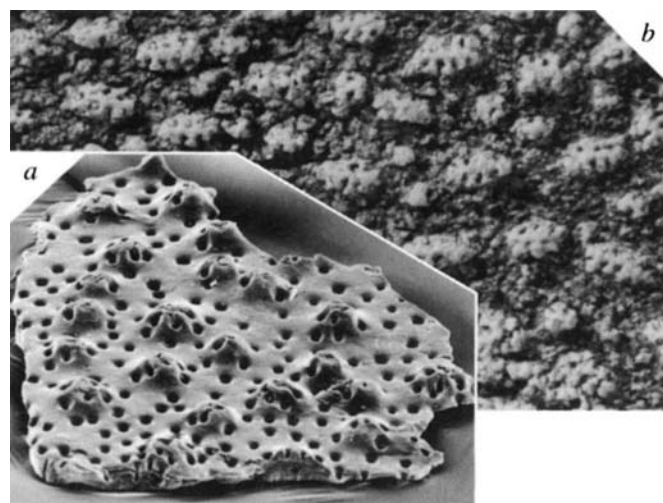


FIG. 1 Oldest known vertebrates from Australia, both characterized by extensive pore-canal systems. a, Oblique view of the new type of dermal armour from the Late Cambrian (CPC33845; magnification, $\times 110$); b, comparable ornament of the Ordovician pteraspidomorph agnathan *Porophoraspis* (CPC33539; magnification, $\times 20$), which differs mainly in the more irregular and elongate tubercles, and much larger size. (CPC, Commonwealth Palaeontological Collection, Australian Geological Survey Organisation, Canberra.)

but shows crescentic elevations and adjacent depressions, similarly spaced to the external tubercles, each with a distinct foramen and ramifying grooves. These we interpret to have transmitted the vascular supply from the corium.

Internal structure was studied using oil of anise, which makes fragments translucent or transparent, and by examining thin sections of one fragment under Nomarski optics. The middle layer reveals a microcalcospheretic organization, which is radially arranged in polygons centred on each vertical canal (Fig. 3a). The tubercles (Fig. 3b) contain irregular canals, smaller than those of the pores, with a main one opening at the visceral surface (Fig. 4). Vertical canals from the pores situated between tubercles do not traverse the basal layer, but are connected by a horizontal system of smaller-diameter canals running at two levels within the tissue.

The general organization of three distinct tissue types, with tubercles and a pore-canal system (Fig. 4), is quite different from fossil arthropod cuticle, which is multilayered, with a finely foliated principal layer accounting for 85% or more of cuticle thickness¹⁸. We interpret the new three-layered skeleton as early vertebrate dermal armour, which is unique in lacking dentine tissue: the tubercles instead are formed from elevations of the superficial (enamel-like) layer. The new evidence impinges on two major issues regarding the evolution of early vertebrates and their hard tissues. First is the long-standing debate concerning the initial condition of the vertebrate exoskeleton. Was it micromeric, formed of many similar small elements (odontodes^{9,10}) like the continuous cover of minute denticles in modern chondrichthyans or Palaeozoic thelodont agnathans? Alternatively, did it first form as large plates or sheets of bone-like tissue (the macromeric condition)? This new Australian discovery, and the Late Cambrian

Anatolepis from North America^{5,6,14}, support the macromeric hypothesis by applying the traditional palaeontological argument that older fossils probably indicate the primitive condition, but this argument is flawed in assuming a complete fossil record. Sampling bias is well illustrated by the previously undetected vertebrate micro-remains from the well studied Harding Sandstone¹⁹, famous for over a century²⁰ because of its vertebrate macrofossils of Ordovician age.

A second issue concerns the sequence of phyletic development of the different hard-tissue types in the early evolution of the vertebrate skeleton, for which the primary evidence is the combination of tissue types observed in early vertebrate fossil remains. If dentine and bone of attachment (odontodes) were the earliest hard tissues to evolve in vertebrates^{11,21}, then identification of dentine would be a reliable indicator of vertebrate affinity. The controversial 'dentine' of euconodonts supports the 'teeth before armour' hypothesis²², but an alternative view is that conodont histology is unique, and derived rather than primitive³. *Anatolepis* can be interpreted to display a primitive condition from which both micro- and macromeric skeletons evolved⁶. An unparsimonious consequence is that this requires first the loss of the lamellar tissue connecting the denticles, then re-formation of macromeric plates using a new tissue type (bone). At least some examples of *Anatolepis* suggest a spongy bone-like middle layer⁴, and the histological evidence may also be seen as consistent with primitive micromery, because of the structural resemblance (homology?) of *Anatolepis* 'denticles' to separate odontodes, which also occur as dermal tubercles in all major groups of macromeric early vertebrates.

The new Australian material demonstrates a unique combina-

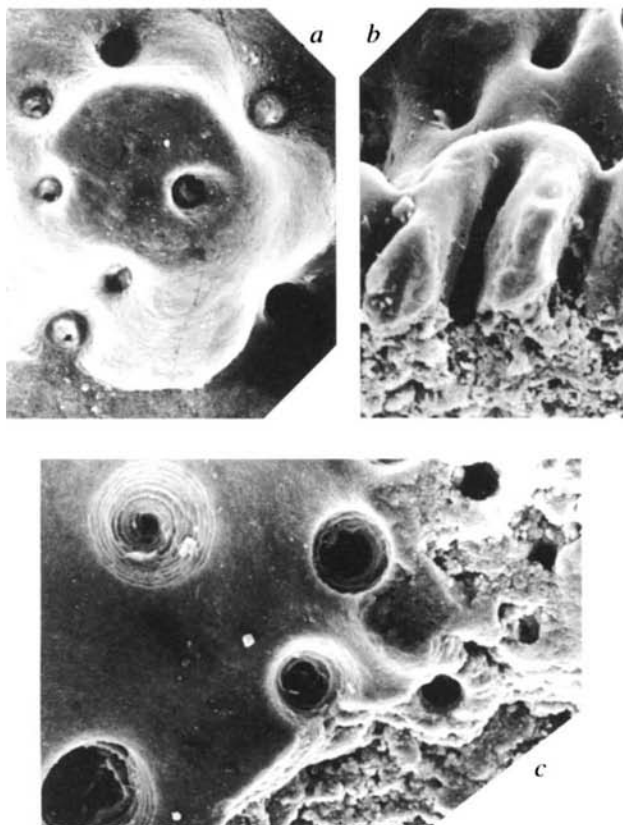


FIG. 2 Late Cambrian vertebrate from Australia. a, Detail of tubercle, CPC33837 (magnification, $\times 1,000$); b, edge of plate showing junction between hypermineralized superficial layer (above) and globular middle layer (below), CPC33838 (magnification, $\times 1,000$); c, detail of pore-canal openings, CPC33830 (magnification, $\times 1,000$).

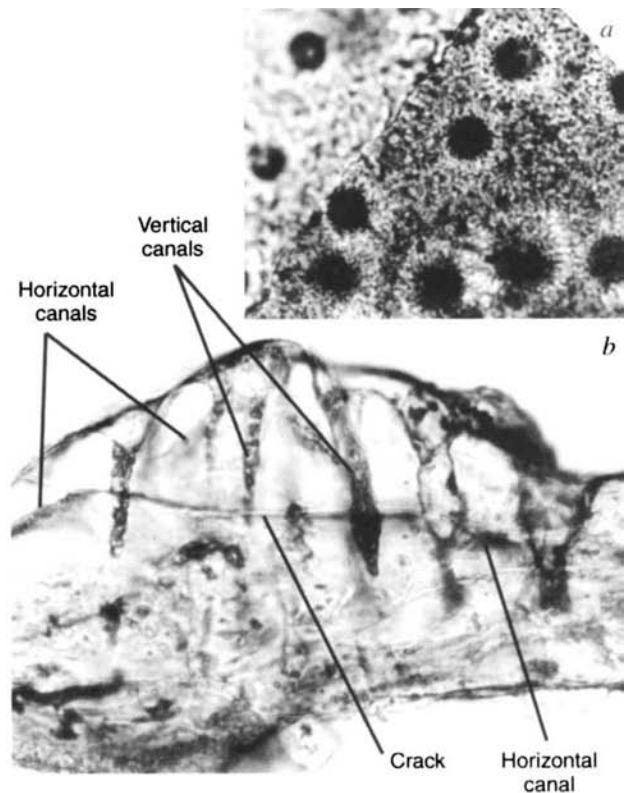


FIG. 3 Late Cambrian vertebrate from Australia. a, Portion of plate under oil of anise, showing radiating spheritic calcification of middle layer around vertical pore canals, CPC33847 (magnification, $\times 180$); b, vertical thin section through a tubercle, CPC33851 (magnification, $\times 750$), showing vertical and horizontal canals (outer and inner systems of the latter are joined by a crack in the tissue).

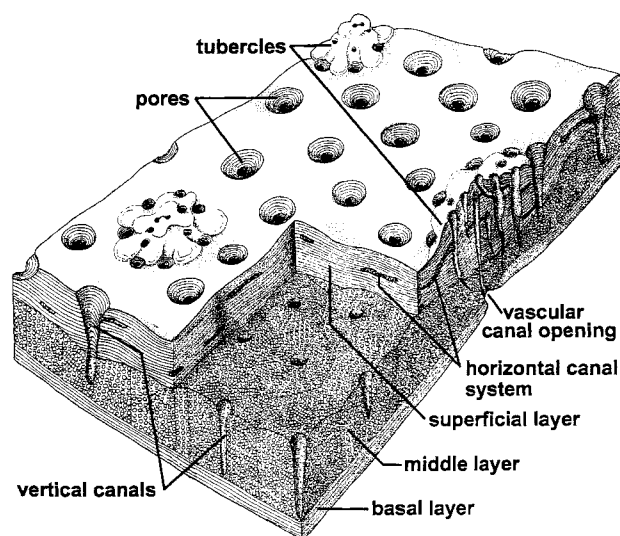


FIG. 4 Late Cambrian vertebrate from Australia. Block diagram summarizing the three-layered structure of the dermal armour with its pore canal and vascular systems.

tion of histological characters, in a sheet-like phosphatic skeleton which lacks odontode units, thereby contradicting the odontode model, but conforming to an alternative view that enamel-like tissues may have been most primitive for vertebrates²³. The highly mineralized surface layer may also have served as a store for calcium and phosphate processed through the epithelium²⁴. Another idea relates the origin of the dermal skeleton to a dentine sensory system, with mineralized sheets fixing their position in the tissue²⁵. The well-developed pore-canal system in our new material suggests an alternative sensory function, perhaps neuromasts as part of a rudimentary lateral line system. The polygonal structure of the middle layer is reminiscent of the tissue cosmine in Devonian sarcopterygian fishes²⁶, in which the flask chambers have been interpreted as sites for electroreceptors²⁷. The pore canals terminating within the middle layer (Fig. 4) could have had a similar function, differing mainly in the more rudimentary system of connection through the horizontal canals. The discovery of a pore-canal system in various agnathan and gnathostome groups suggests that it formed an essential part of the primitive laterosensory system of vertebrates^{1,28}. In contrast, the fragments of *Anatolepis* lack structures interpretable as parts of a laterosensory system, the 'pore canals' recently described⁶ as traversing the laminar tissue between the odontodes being quite different in size and structure, with no evidence of interlinking by horizontal canals. This laminar tissue in *Anatolepis* has been compared with 'hyaline' in some recent teleost scales, but this is a mesodermally derived tissue formed by osteoblast-like cells, in which the only spaces are for attachment fibre bundles.

The discovery in the Late Cambrian of two completely different types of phosphatic dermal armour suggests experimentation in initial evolution of the vertebrate skeleton, which must have occurred much earlier than previously assumed. This is consistent with the prediction that major chordate clades (tunicates, craniates) had already evolved in the main burst of the Cambrian explosion²⁹. This new material combines a pore-canal system with enamel-like hard tissues, implying that these may have preceded the formation of bone and dentine, with both euconodonts^{3,12,13,22} and *Anatolepis*^{6,14} representing divergent specializations within the early diversification of vertebrate hard tissues. □

5. Repetski, J. E. *Science* **200**, 529–531 (1978).
6. Smith, M. P., Sansom, I. J. & Repetski, J. E. *Nature* **380**, 702–704 (1996).
7. Ørving, T. in *Structural and Chemical Organization of Teeth* (ed. Miles, A. E. W.) 45–110 (Academic, London, 1967).
8. Ørving, T. in *Current Problems in Lower Vertebrate Phylogeny* (ed. Ørving, T.) 374–397 (Almqvist & Wiksell, Stockholm, 1968).
9. Ørving, T. in *Problems in Vertebrate Evolution* (eds Andrews, S. M., Miles, R. S. & Walker, A. D.) 53–75 (Academic, London, 1977).
10. Reif, W.-E. *Evolut. Biol.* **15**, 287–368 (1982).
11. Smith, M. M. & Hall, B. K. *Biol. Rev.* **65**, 277–373 (1990).
12. Sansom, I. J., Smith, M. P., Armstrong, H. A. & Smith, M. M. *Science* **256**, 1308–1311 (1992).
13. Sansom, I. J., Smith, M. P. & Smith, M. M. *Nature* **368**, 591 (1994).
14. Smith, M. P. & Sansom, I. J. *Geobios. Mém. Spéc.* **19**, 61–64 (1995).
15. Nicoll, R. S. *AGSO J. Aust. Geol. & Geophys.* **15**, 367–379 (1994).
16. Ritchie, A. & Gilbert-Tomlinson, J. *Alcheringa* **1**, 351–368 (1977).
17. Young, G. C. & Laurie, J. R. (eds) *An Australian Phanerozoic Timescale* (Oxford Univ. Press, Melbourne, 1996).
18. Dalingwater, J. E., Hutchinson, S. J., Mutvei, H. & Siveter, D. J. *Palaeontographica Abt. A* **229**, 37–49 (1993).
19. Sansom, I. J., Smith, M. M. & Smith, M. P. *Nature* **379**, 628–630 (1996).
20. Walcott, C. D. *Geol. Soc. Am. Bull.* **3**, 153–172 (1892).
21. Smith, M. M. *Science* **251**, 301–303 (1991).
22. Sansom, I. J., Smith, M. P. & Smith, M. M. *Modern Geol.* (in the press).
23. Maisey, J. G. *Evol. Biol.* **22**, 1–36 (1988).
24. Halstead, L. B. *Vertebrate Hard Tissues* (Wykeham, London and Winchester, 1974).
25. Gans, C. *Biol. Rev.* **64**, 221–268 (1989).
26. Gross, W. *Kungl. Svensk VetenskapsAkad. Handl. 4 Ser.* **5**, 1–140 (1956).
27. Thomson, K. S. in *Problems in Vertebrate Evolution* (eds Andrews, S. M., Miles, R. S. & Walker, A. D.) 247–270 (Academic, London, 1977).
28. Northcutt, R. G. & Gans, C. *Quart. Rev. Biol.* **58**, 1–28 (1983).
29. Chen, J.-Y., Dzik, J., Edgecombe, G. D., Ramsköld, L. & Zhou, G.-Q. *Nature* **377**, 720–722 (1995).

ACKNOWLEDGEMENTS. We thank R. S. Nicoll and J. H. Shergold for the material, with information on occurrence and age; R. W. Brown, C. Thun and A. Kelman for assistance with preparation, SEM photography and figures; I. Sansom for preparing the thin section; R. A. Fortey and W. D. I. Rolfe for advice on arthropod cuticle; and R. S. Nicoll, J. H. Shergold, C. B. Foster, P. Janvier and P. Forey for comments on the original manuscript. V.N.K.-T. acknowledges S. Turner's support for a visit to Australia in 1994 under IGCP project 328 (Palaeozoic microvertebrates), to which this work contributes. M.M.S. acknowledges an NERC project grant and the Royal Society for an exchange visit for V.N.K.-T.

CORRESPONDENCE and requests for materials should be addressed to G.C.Y. (e-mail: gyoung@agso.gov.au).

A differential neural response in the human amygdala to fearful and happy facial expressions

J. S. Morris*, C. D. Frith*, D. I. Perrett†, D. Rowland‡, A. W. Young‡, A. J. Calder‡ & R. J. Dolan*§

* Wellcome Department of Cognitive Neurology, Queen Square, London WC1N 3BG, UK

† School of Psychology, University of St Andrews, Fife KY16 9JU, UK

‡ MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, UK

§ Royal Free and University College Hospitals School of Medicine, Rowland Hill Street, London NW3 2DF, UK

THE amygdala is thought to play a crucial role in emotional and social behaviour¹. Animal studies implicate the amygdala in both fear conditioning² and face perception³. In humans, lesions of the amygdala can lead to selective deficits in the recognition of fearful facial expressions^{4,5} and impaired fear conditioning^{6,7}, and direct electrical stimulation evokes fearful emotional responses⁸. Here we report direct *in vivo* evidence of a differential neural response in the human amygdala to facial expressions of fear and happiness. Positron-emission tomography (PET) measures of neural activity were acquired while subjects viewed photographs of fearful or happy faces, varying systematically in emotional intensity. The neuronal response in the left amygdala was significantly greater to fearful as opposed to happy expressions. Furthermore, this response showed a significant interaction with the intensity of emotion (increasing with increasing fearfulness, decreasing with increasing happiness). The findings provide direct evidence that the human amygdala is engaged in processing the emotional salience of faces, with a specificity of response to fearful facial expressions.

Received 2 April; accepted 30 August 1996.

1. Young, G. C. *J. Vert. Paleontol.* (in the press).

2. Forey, P. L. & Janvier, P. *Am. Scientist* **82**, 554–565 (1994).

3. Janvier, P. *Nature* **374**, 761–762 (1995).

4. Bockellie, T. & Fortey, R. A. *Nature* **260**, 36–38 (1976).

Facial expressions are a mechanism through which internal emotional states and intentions become available as external signals, consequently the face is vital in social cognition^{1,9,10}. The primate amygdala is thought to be involved in processing facial expression and in controlling emotional and social behaviour^{1,2}. Direct evidence for its role in humans is limited to neuropsychological studies of subjects with rare selective amygdala lesions^{4,5}, and electrophysiological studies of epileptic patients⁸. Using normal volunteer subjects, we tested whether the human amygdala is critical in processing fearful expressions, as suggested by the lesion studies^{4,5}. We measured regional cerebral blood flow (rCBF) in five subjects while they viewed grey-scale images of emotionally expressive faces taken from a standard set of pictures of facial affect¹¹. The faces depicted either happy or fearful expressions, and for each emotional category and individual face a range of six levels of intensity was produced by computer graphical manipulation¹² (Fig. 1). For each subject, separate scans were acquired for each intensity level in a 2×6 (category $\times$ intensity) factorial experimental design. For each presented face, subjects were simply required to make a gender classification (male or female) by pressing left or right response buttons. No explicit recognition or categorization of emotional expression was required during the scans, and post-scan debriefing confirmed that subjects were not aware that the implicit emotional variable was crucial in the experimental design.

The rCBF pattern produced by the presentation of fearful faces (contrasted with happy) showed the left amygdala and left periamygdaloid cortex as the most significant areas of activation ($P < 0.05$, corrected). There was no activation of the right amygdala, even at low thresholds. The statistical parametric map (SPM) of this contrast and the associated mean adjusted rCBF values for the happy and fearful conditions at the maximal point of activation are shown in Fig. 2. Other areas, that would have been judged significant if they had been predicted beforehand, included the left cerebellum, left cingulate gyrus and right superior frontal gyrus ($P < 0.001$, uncorrected). A contrast of happy with fearful expressions, on the other hand, was associated with activations in the right medial temporal gyrus, right putamen, left superior parietal lobe, and left calcarine sulcus ($P < 0.001$, uncorrected). The coordinates and Z scores for these activations are given in Table 1.

TABLE 1 Regions of activity in response to emotional expression contrasts

Fearful–happy		
Area	Coordinates (x, y, z)	Z score
Left periamygdaloid cortex (BA34)	−14, −8, −20	4.26
Left amygdala	−18, −6, −16	3.86
Left cerebellum	−42, −68, −20	3.57
Right superior frontal gyrus (BA6)	22, 4, 64	3.22
Left cingulate gyrus (BA23)	−10, 28, 16	3.18
Happy–fearful		
Right medial temporal gyrus (BA21)	54, 4, −20	3.55
Right putamen	22, −4, 12	3.48
Left superior parietal lobule (BA5)	−28, −40, 60	3.41
Left superior parietal lobule (BA7)	−24, −72, 44	3.12
Left calcarine sulcus (BA17)	−10, −92, 4	3.06

Regions selectively activated when the different emotional conditions are contrasted. The contrasts were fearful relative to happy expressions (top) and happy relative to fearful expressions (bottom). Coordinates of the maximal point of activation, Brodmann areas (BA), and the associated Z values are shown. The activations in all regions are significant at $P < 0.001$ (uncorrected). In the amygdala, the only area predicted to show a response, this is equivalent to a significance level of $P < 0.05$, corrected for multiple spatial comparisons in a $2 \times 2 \times 2$ cm search region³⁰. The activations in the other brain regions would have had the same corrected level of significance if they had also been predicted beforehand. All P values are one-tailed.

The interaction between the neural responses to emotional category and emotional intensity was examined by applying orthogonal contrasts. Thus, the voxels that showed a significant response in the contrast of fearful versus happy faces (as in Fig. 2) were subjected to a second orthogonal contrast to determine the presence of a differential response to varying emotional intensity (increasing activation with increasing fearfulness and decreasing activation with increasing happiness). The left amygdala was the only significant area in this interaction (Fig. 3). The rCBF response in the left amygdala to changes in the intensity of facial expression shows monotonic increases in left amygdala activation from the most happy condition to the most fearful (Fig. 3).

Our results provide direct evidence that the human amygdala is involved in a neural response to fearful facial expressions. The nature of our experimental design, in which faces were classified by gender and not by expression, suggests that amygdala activation is not contingent upon explicit processing of facial expression, a finding in-keeping with electrophysiological data¹³. The differential response within the amygdala, to fearful as opposed to happy faces, complements psychological data based on multidimensional scaling, indicating that these facial emotions are at opposite ends of a spectrum of similarity⁴. The unilateral response in the left amygdala parallels findings in a study of procaine-induced emotional states¹⁴ where left (but not right) amygdala rCBF correlated positively with fear and negatively with euphoria. It is also consistent with a study of patients with unilateral amygdala damage which found that ratings of emotional intensity in facial expressions were significantly lower with left amygdala lesions compared with right¹⁵. Bilateral amygdala lesions may not always be sufficient to impair recognition of fearful expressions¹⁶, indicating that other neural systems can perform this role. The additional activations associated with fearful faces, in the cerebellum, frontal lobe and cingulate gyrus, which were not predicted, may indicate brain areas, outside the amygdala, which are involved in responding to fearful expressions.

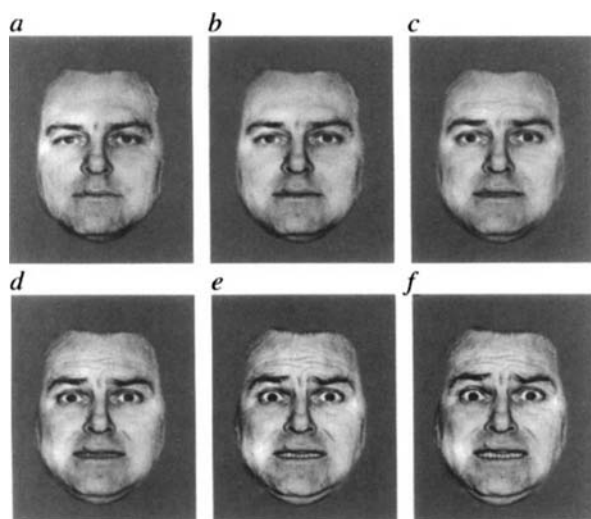


FIG. 1 Faces *a* and *e* are prototypical neutral and fearful expressions. Faces *b*–*d* are interpolated between these prototypes, using computer morphing procedures¹² to shift the shape and pigmentation of the neutral prototype towards the fear prototype; *b* involves 25% fear (and 75% neutral), *c* 50% fear (and 50% neutral), *d* 75% fear (and 25% neutral). Face *f* is an enhanced 125% fear expression, created by shifting the shape of the fear prototype 25% away from neutral (increasing by 25% any differences from neutral).

Neurophysiological studies³ in the macaque and neuroimaging experiments^{17,18} in humans suggest neural segregation of the processing of facial identity and facial emotion. The human studies used explicit tests of emotional recognition and discrimination in which the emotional category and intensity of facial expression were not variables of interest. In our experiment the emotional variable was implicit, and the parametric factorial design provided a means to determine a differential response to both emotional category and intensity. These differences in experimental design may explain why our study of facial emotion did not observe activation in the areas of prefrontal and cingulate cortices reported in the earlier PET experiments. It is possible that there was a common response in these areas in the fearful and happy conditions resulting in no net activation when these emotions are contrasted directly. Alternatively, these frontal areas may be involved in the explicit rather than implicit processing of facial expression, and therefore have a primary role in attentionally related processes.

The primate amygdala receives substantial inputs from temporal visual-association¹⁹⁻²¹, and contains neurons that respond selectively to individual faces^{5,22,23}. The amygdala also has a strong anatomical link with the autonomic nervous system²⁴. This connectivity, coupled with the selectivity of its response, suggests that the amygdala is appropriately placed in relation to sensory and autonomic systems to enable integrated responses to the emotional significance of complex stimuli. Integrated responses to threat or danger, without the necessity of higher level processing, can be mediated by the amygdala^{1,2}. There is evidence in humans that amygdala damage impairs the formation of conditioned autonomic responses to aversive stimuli^{8,9}. We suggest that perceiving an expression of fear in a conspecific may trigger an automatic response to potential danger that accounts for the observed amygdala activation in response to fearful faces.

Ethnological studies suggest that facial expressions of emotion are innate, automatic, and of critical importance in social behaviour^{9,10}. In the nonhuman primates there is evidence that the amygdala is an important component of the neural system involved in social behaviour. Monkeys that have had their amygdala ablated are tame^{25,26}, and no longer make appropriate

responses to signals of danger or threat¹. Radiotelemetry recordings of amygdala activity in monkeys during social interactions show the highest responses to ambiguous or threatening situations (for example, threat face display), and the lowest to tension-lowering behaviours (such as grooming and huddling)²⁷. Our observation of a differential response in the left human

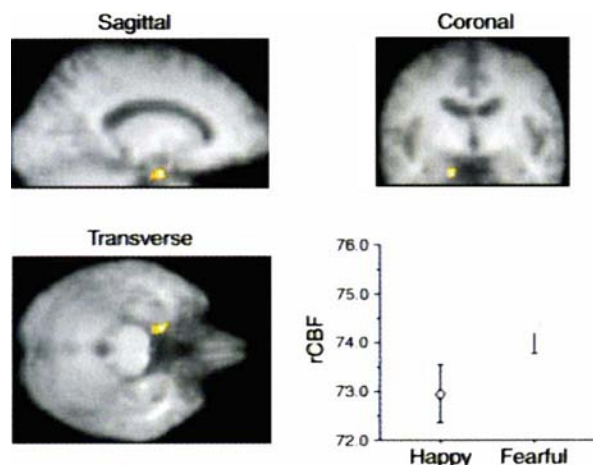


FIG. 2 A statistical parametric map (SPM) showing activation of the left amygdala, together with a graphical display of the mean adjusted rCBF values. An uncorrected P value of 0.01 was used as the threshold for the contrast of the fearful with the happy conditions. Views of the brain are shown for orthogonal slices at the pixel of maximal activation within the left amygdala ($x = -18$, $y = -6$, $z = -16$). The significant area of activation is displayed on the mean MRI image produced from the co-registered structural MRIs from all five subjects. In the graph, unweighted means (± 2 s.e.m.) of the rCBF values at the pixel $x = -18$, $y = -6$, $z = -16$ are shown for the five fearful and five happy conditions. The difference between the means is significant ($P = 0.0014$) on an unpaired t -test.

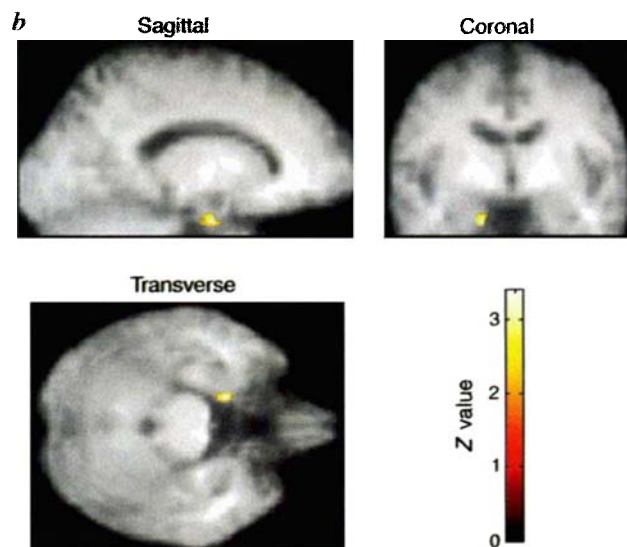
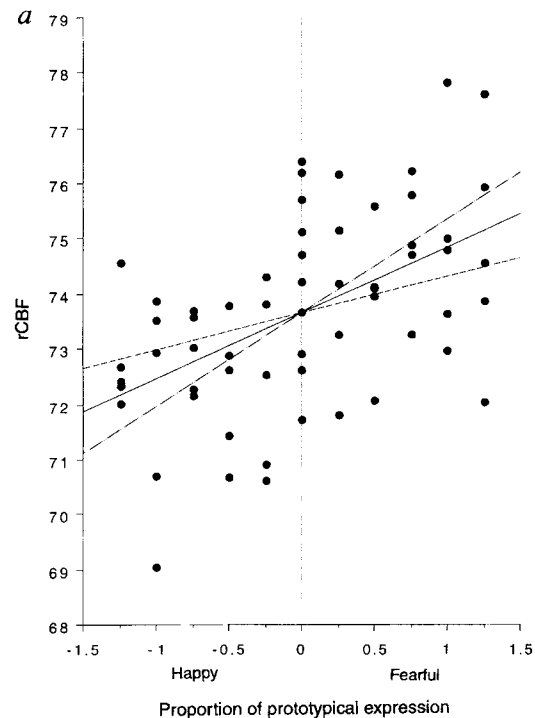


FIG. 3 a, rCBF values and b, statistical parametric maps (SPM) showing activation of the left amygdala in the interaction of emotional category and intensity. The SPM is the result of two orthogonal contrasts (see text). An uncorrected P value of 0.03 was used for both contrasts giving a resultant value of 0.0009 for the interaction. The views of the brain are the same as in Fig. 2. The graph plots the rCBF values in $\text{ml dl}^{-1} \text{min}^{-1}$ for all conditions and all subjects at $x = -18$, $y = -6$, $z = -16$ (the pixel of maximal activation). The x axis represents the proportion of the prototypical expression in the face stimuli, with fearful being positive (100% = 1), and happy negative (100% = -1). A regression line has been fitted to the data (with broken lines representing 95% confidence intervals for the gradient of the slope. Correlation coefficient r , 0.514.

amygdala to faces that embody either fearful or happy expressions indicates that it may play a similar regulatory role in human social behaviour. □

Methods

Subjects. Four male subjects and one female subject (mean age 42.8 years) took part in the study (approved by the local hospital ethics committee and ARSAC(UK)). All subjects were healthy, with no past history of psychiatric or neurological illness, and were not on any medication. A sixth post-menopausal female subject (age 62 years) was scanned but she was later found to be receiving hormone replacement therapy and was therefore excluded. Inclusion of the data from this subject in the analysis still produced significant activation of the left amygdala ($Z = 3.35$).

PET scan acquisition and analysis. Scans of the distribution of $H_2^{15}O$ were obtained using a Siemens-CPS ECAT EXACT HR⁺ PET Scanner operated in high sensitivity three-dimensional mode. Subjects received a total of 350 MBq of $H_2^{15}O$ over 20 s through a forearm cannula. Images were reconstructed into 63 planes, using a Hann filter, resulting in a 6.4 mm transaxial and 5.7 mm axial resolution (full-width at half-maximum). The data were analysed with statistical parametric mapping (SPM 95 software, Wellcome Department of Cognitive Neurology, London) implemented in Matlab (Mathworks, Sherborn, MA). After initial realignment, mean PET images from each subject were scalp-edited and used as a template to edit all 12 individual PET images. Structural magnetic-resonance images (MRIs) from each subject were co-registered into the same space. The scans were then transformed into a standard stereotactic space²⁸. The scans were smoothed using a gaussian filter set at 12 mm full-width at half-maximum. The rCBF measurements were adjusted to a global mean of 50 ml dl⁻¹ min⁻¹. A blocked (by subject) ANCOVA model was fitted to the data at each voxel, with a condition effect for each level of emotional intensity, and global CBF as a confounding covariate. Predetermined contrasts of the condition effects at each voxel were assessed using the usual *t*-statistic, giving a statistic image for each contrast. The method of SPM data analysis is described in detail in refs 28 and 29.

Experimental design. During each scan, 10 photographs of faces were presented, one at a time, on a computer monitor screen. Each presentation lasted for 3 s, followed by a 2-s interval in which the screen was blank. The 10 faces were of different individuals (5 males and 5 females), but all had the same category and intensity of emotional expression. The faces of the same 10 individuals were used in all 12 scans, in a randomized order. The emotional category and intensity of the faces were varied systematically across scans. The order of presentation of happy and fearful conditions was counterbalanced across subjects. The six different intensity levels were given in a counter-balanced order within and across subjects. In the gender-classification task during scanning, all five subjects identified correctly 90–100% of the time.

Behavioural tests. The validity of the different emotional categories and 'morphed' intensity levels was confirmed using rating and discrimination tests performed on the same subjects, after scanning. In classifying the category (fearful, happy, neutral or 'other'), 96.5% of responses were correct. Subjects then rated the intensity of expression of the face on a 7-point scale. These ratings correlated well with the proportion of the fearful or happy prototype in the morphed face (correlation coefficients: $r = 0.772$ for fearful, $r = 0.826$ for happy). In a separate discrimination test of emotional intensity, different faces were presented in pairs, and subjects selected the more intense expression. Again, there was a close agreement between perceived and 'morphed' intensities: 69.4% of ratings agreed for pairs differing by 25% in their percentage of prototype, 83.3% agreed for pairs differing by 50%, and there was 100% agreement for pairs differing by >50%.

Received 4 June; accepted 11 September 1996.

- Kling, A. S. & Brothers, L. A. In *The Amygdala: Neurobiological Aspects of Emotion, Memory and Mental Dysfunction* (ed. Aggleton, J. P.) 353–377 (Wiley-Liss, New York, 1992).
- LeDoux, J. E. *Behav. Brain Res.* **58**, 69–79 (1993).
- Rolls, E. T. In *The Amygdala: Neurobiological Aspects of Emotion, Memory and Mental Dysfunction* (ed. Aggleton, J. P.) 143–167 (Wiley-Liss, New York, 1992).
- Adolphs, R., Tranel, D., Damasio, H. & Damasio, A. *Nature* **372**, 669–672 (1994).
- Calder, A. J. et al. *Cogn. Neuropsychol.* **13**, 699–745 (1996).
- Bechara, A. et al. *Science* **269**, 1115–1118 (1995).
- LaBar, K. S., LeDoux, J. E., Spencer, D. D. & Phelps, E. A. *J. Neurosci.* **15**, 6846–6855 (1995).
- Halgren, E., Babb, T. C., Rausch, R. & Crandall, P. H. *Brain* **101**, 83–117 (1978).
- Darwin, C. *The Expression of the Emotions in Man and Animals* (Univ. Chicago Press, 1965).
- Ekman, P. *Emotion in the Human Face* (Cambridge Univ. Press, Cambridge, 1982).
- Ekman, P. & Friesen, W. V. *Pictures of Facial Affect* (Consulting Psychologists, Palo Alto, 1976).
- Perrett, D. I., May, K. A. & Yoshikawa, S. *Nature* **368**, 239–242 (1994).
- Halgren, E. In *The Amygdala: Neurobiological Aspects of Emotion, Memory and Mental Dysfunction* (ed. Aggleton, J. P.) 190–228 (Wiley-Liss, New York, 1992).
- Ketter, T. A. et al. *Arch. Gen. Psychiatry* **53**, 59–69 (1996).
- Adolphs, R., Tranel, D., Damasio, H. & Damasio, A. *J. Neurosci.* **15**, 5879–5891 (1995).
- Hamann, S. B. et al. *Nature* **379**, 497 (1996).
- George, M. S. et al. *J. Neuropsychiatry Clin. Neurosci.* **5**, 384–394 (1993).
- Sergent, J., Ohta, S., MacDonald, B. & Zuck, E. *Vis. Cogn.* **1**, 349–369 (1994).
- Herzog, A. G. & Van Hoesen, G. W. *Brain Res.* **115**, 57–69 (1976).
- Iwai, E. & Yukie, M. *J. Comp. Neurol.* **261**, 362–387 (1987).

- Aggleton, J. P., Burton, M. J. & Passingham, R. E. *Brain Res.* **190**, 347–368 (1980).
- Perrett, D. I. et al. *Hum. Neurobiol.* **3**, 197–208 (1984).
- Perrett, D. I. et al. *J. Exp. Biol.* **146**, 87–114 (1989).
- LeDoux, J. E., Iwata, J., Cicchetti, P. & Reis, D. J. *J. Neurosci.* **10**, 1062–1069 (1988).
- Kluver, H. & Bucy, P. C. *Arch. Neurol. Psychiatry* **42**, 979–1000 (1939).
- Weiskrantz, L. J. *Comp. Physiol. Psychol.* **4**, 381–391 (1956).
- Kling, A. S., Steklis, H. D. & Deutsch, S. *Exp. Neurol.* **66**, 88–96 (1979).
- Friston, K. J. et al. *Hum. Brain Mapp.* **2**, 189–210 (1995).
- Friston, K. J. et al. *Hum. Brain Mapp.* **3**, 165–189 (1995).
- Worsley, K. J. et al. *Hum. Brain Mapp.* (in the press).

ACKNOWLEDGEMENTS. We thank K. Friston for conceptual advice. J.S.M., C.D.F. and R.J.D. are supported by the Wellcome Trust.

CORRESPONDENCE and requests for materials should be addressed to R.J.D. (e-mail: rdolan@fil.ion.ucl.ac.uk).

Precisely correlated firing in cells of the lateral geniculate nucleus

Jose-Manuel Alonso, W. Martin Usrey* & R. Clay Reid*

Laboratory of Neurobiology, The Rockefeller University, New York, New York 10021, USA

SIMPLE cells within layer IV of the cat primary visual cortex are selective for lines of a specific orientation. It has been proposed that their receptive-field properties are established by the pattern of connections that they receive from the lateral geniculate nucleus (LGN) of the thalamus^{1–5}. Thalamic inputs, however, represent only a small proportion of the synapses made onto simple cells^{6–8}, and others have argued that corticocortical connections are likely to be important in shaping simple-cell response properties^{9–11}. Here we describe a mechanism that might be involved in selectively strengthening the effect of thalamic inputs. We show that neighbouring geniculate neurons with overlapping receptive fields of the same type (*on*-centre or *off*-centre) often fire spikes that are synchronized to within 1 millisecond. Moreover, these neurons often project to a common cortical target neuron where synchronous spikes are more effective in evoking a postsynaptic response. We propose that precisely correlated firing within a group of geniculate neurons could serve to reinforce the thalamic input to cortical simple cells.

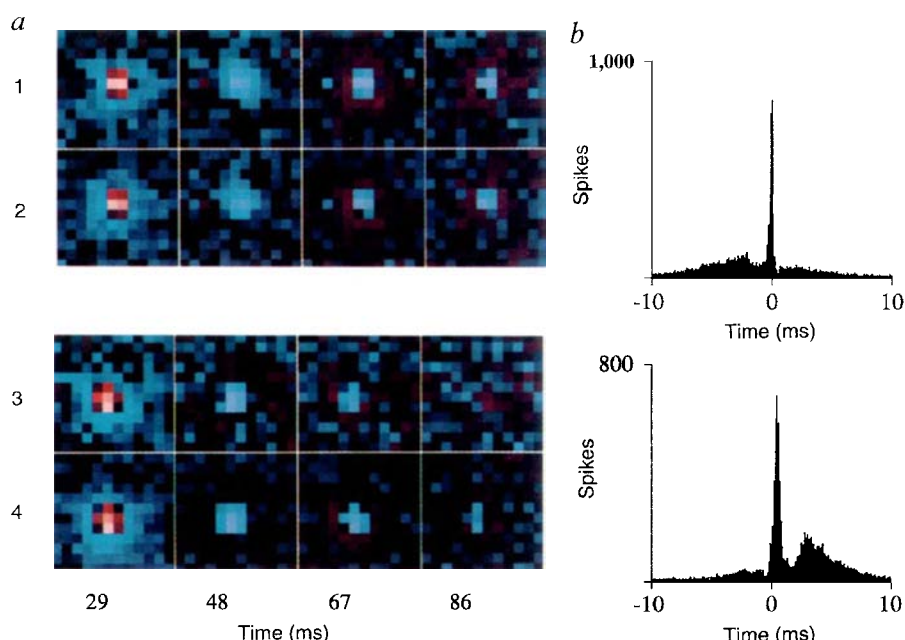
The projections from the retina to the thalamus are both divergent and convergent. Individual retinal ganglion cells diverge to connect to at least four different geniculate neurons¹². Conversely, geniculate neurons often receive convergent input from two or more ganglion cells¹³. Although slow correlations have been noted between geniculate neurons^{14–16}, little attention has been paid to a faster synchrony that may be caused by common input from divergent retinal afferents. Here we have found that extremely fast correlations do exist, and propose a role for them in the transmission of information from thalamus to cortex.

In simultaneous recordings from pairs of closely positioned geniculate cells (electrodes 100–400 μ m apart), we occasionally observed firing patterns that were tightly correlated at the 1-ms timescale. These fast correlations were narrower and often stronger than other correlations previously studied in the visual system: between retinal ganglion cells^{17–19}, between geniculate cells with non-overlapping receptive fields^{15,16}, between geniculate and cortical cells^{3,4} or between cortical cells^{20–22}.

The strongest fast correlations were between geniculate cells with very similar receptive fields (same position, sign, size and timing). Figure 1a shows the receptive fields of two pairs of *on*-centre geniculate X cells calculated by reverse correlation^{23,24}. For

* Present address: Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115, USA.

FIG. 1 Two examples of strong, tight correlations between X cells with very similar receptive fields at different delays between stimulus and response. Red, on responses. Blue, off responses. The brightest colours correspond to the strongest responses. Scale: 0.4 deg per pixel. *a*, Receptive fields. Pair 1–2: two on-centre X cells with well overlapped receptive fields of similar size and timing. There is a small difference in the position of the strongest pixels at the centre of the receptive fields. Pair 3–4: two other on-centre X cells with almost identical receptive fields. There is only a small difference in the time course of the visual responses; the response in cell 3 is less sustained than in cell 4 (compare the relative strength of the receptive fields at 67 and 86 ms). *b*, Cross-correlations. Pair 1–2: half-width, 0.5 ms; strength, 25%. Pair 3–4: half-width, 1 ms; strength, 43%. The two neurons in each pair were recorded with different electrodes.

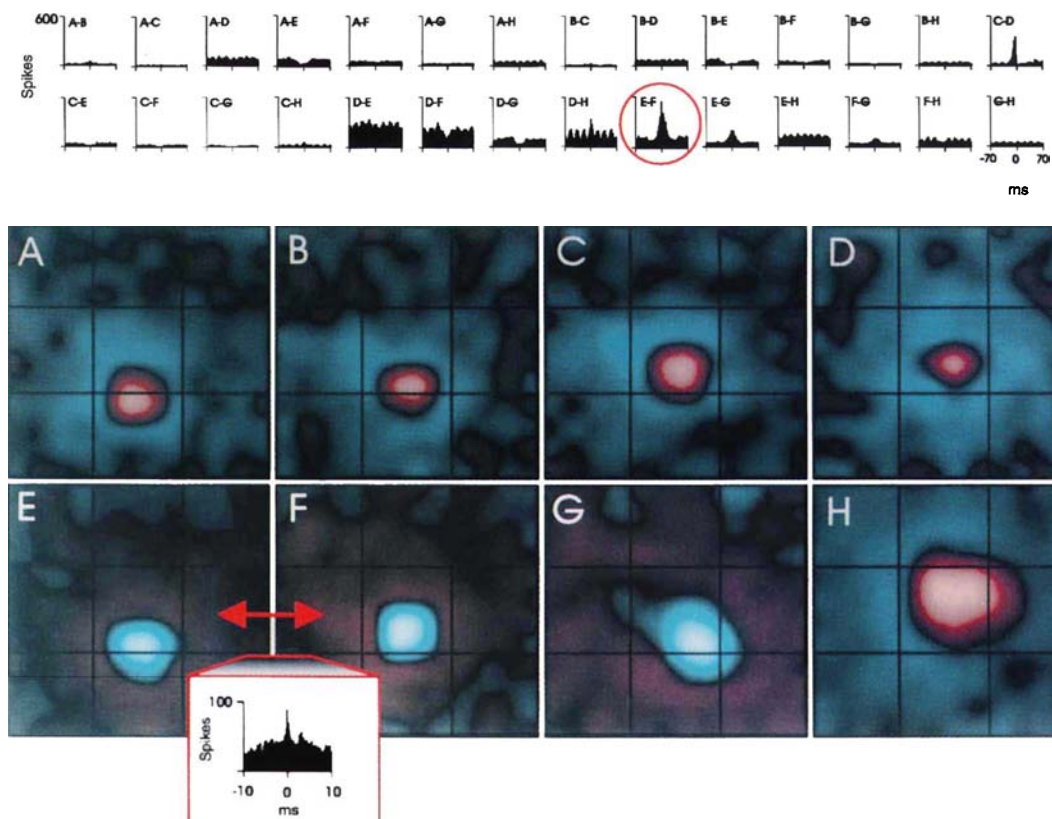


each cell, the time course of the visual response is shown by a series of receptive-field maps calculated for different delays between stimulus and response. The visual responses of each pair of cells (1–2, 3–4) were very similar. When the firing patterns of the cells in each pair were studied by cross-correlation analysis (Fig. 1*b*), we observed very narrow and strong peaks near zero, indicating the neurons often produced spikes within 1 ms of each other (strength of correlation 1–2, 25%; 3–4, 43%).

In addition to the strong correlations seen between cells with

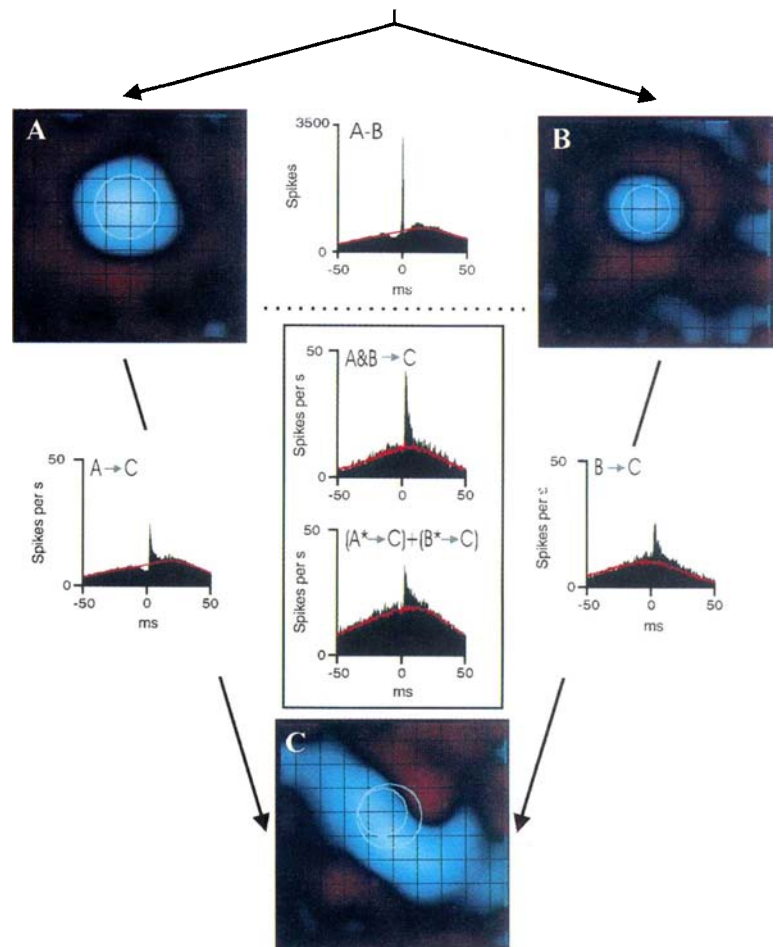
nearly identical receptive fields, we occasionally observed weaker correlations between cells with partially overlapped receptive fields. An example of one of these weaker correlations is illustrated in Fig. 2. Shown are the receptive fields of 7 geniculate cells and one synaptic potential, from a retinal input, recorded simultaneously with 5 electrodes. Of the 21 cross-correlations between geniculate cells, only one narrow peak was observed (red circle and inset). This correlation was much weaker (3.3%) than the correlations between cells with nearly identical receptive fields (Fig. 1).

FIG. 2 One example of a weak tight correlation between two X cells with partially overlapped receptive fields (cells E and F). Seven geniculate neurons (A,B,C,E,F,G,H) and a synaptic potential from a retinal input (D) were recorded simultaneously with five different electrodes. The top of the figure shows the 28 possible cross correlations calculated within a time window of 70 ms. E–F (red circle) is the only example of two tightly correlated LGN cells (half-width, 0.5 msec; strength, 3.3%). In the 70 ms correlation, the thin peak is superimposed over the stimulus-dependent correlation. The 10 ms cross-correlation (inset) between cells E and F shows the narrow peak in greater detail. Notice that the strength of this correlation is much weaker than the one shown in Fig. 1. C–D is a correlation between a synaptic potential (D) and an LGN cell (C). The oscillations seen in some of the correlations are stimulus dependent (the white-noise stimulus was



updated at 52 Hz). The receptive fields have been smoothed by one half pixel. Black squares correspond to 4×4 stimulus pixels, or 1.6° .

FIG. 3 Two strongly correlated off-centre X cells connected to the same simple cell. Receptive fields and cross-correlations: Panels A and B show the receptive fields of two off-centre X cells. C is the receptive field of a simple cell. The receptive fields have been smoothed by one half pixel. Black squares correspond to one stimulus pixel, or 0.4° . The two geniculate cells show strong, tightly correlated firing (A-B correlation) which indicates that they probably receive common input from the same ganglion cell (schematically represented by arrows at the top). The cross-correlations between the simple cell and each geniculate cell (A $\rightarrow$ C, B $\rightarrow$ C) show peaks displaced 2–3 ms to the right of zero which indicate that the simple cell receives monosynaptic input from each geniculate cell^{3,4} (schematically represented by arrows on the sides). The red lines are the stimulus-dependent correlations (shuffle correlations³⁰). Unlike the intrageniculate correlations, the geniculocortical correlations have a clear presynaptic cell. Genulocortical correlations, therefore, have been normalized by the number of presynaptic spikes and then expressed in units of spikes per second for the cortical cell. Efficacy of the geniculocortical connection: The spike trains of neurons A and B were divided into two categories: those that occurred within 1.0 ms of each other (A&B) and those of either neuron that did not (A* and B*). The influence of the simultaneous spikes (A&B) was compared to that of the non-simultaneous spikes by adding together the correlograms from A* $\rightarrow$ C and B* $\rightarrow$ C (middle panels). Efficacy was calculated as the percentage of thalamic spikes followed by a cortical spike after subtracting the shuffle correlation^{3,4}. The summed efficacy of the non-simultaneous spikes (A* and B*) was $3.12 \pm 0.30\%$ (efficacy of A* = $1.71 \pm 0.07\%$; efficacy of B* = $1.41 \pm 0.22\%$) whereas that of the simultaneous spikes (A&B) was $5.28 \pm 0.23\%$. Each of these values may be slight underestimates of the actual efficacies. This is because the baseline subtraction, or shuffle correction, itself includes a contribution from the geniculate neuron under study. This potential source of error, however, is relatively small compared with the 69% difference in calculated efficacies. As these neurons were studied over a long period, there were sufficient spikes in all categories to study their influence on C (A: 64,779; B: 16,784; A*: 58,054; B*: 10,059; A&B: 13,450; C: 8,635).



We recorded from 282 pairs of geniculate cells with some degree of overlap between their receptive field centres or surrounds (Table 1). All of the cell pairs with nearly identical receptive fields were correlated on the ms timescale; further, these correlations were very strong (average, 28%). Receptive fields that differed in any respect (position, size, sign or timing) were less likely to be correlated and the correlations were much weaker.

Several lines of evidence suggest that the strong, tight correlations shown in Fig. 1 are the result of inputs from a common retinal ganglion cell. Strong correlations were seen only between geniculate cells with very similar receptive fields (Table 1). Additionally, in two experiments we observed a common synaptic potential or a T-waveform (spike from the retinal axon) preceding the spikes of both correlated cells. The rather low percentage of tightly correlated cells (15%) fits well with the anatomical data, which indicates that only 10% of neighbouring geniculate cells receive input from any given ganglion cell¹².

The precisely correlated firing of geniculate neurons could be important for the development of the thalamocortical projection²⁵ or may provide the substrate for neuronal coding strategies based on precise timing^{19,26}. A related possibility is that synchronous inputs from several thalamic cells could interact to increase their effectiveness in driving a common cortical target.

To test the hypothesis that simultaneous thalamic spikes are more effective than non-simultaneous spikes, we examined the interactions between several geniculate cells connected to a common simple cell. Figure 3 shows an example of two precisely correlated, off-centre X cells (A, B in Fig. 3) connected to the same simple cell (C). The correlogram at the top of the figure (A-B) shows a strong, fast correlation between the two X cells (as in Fig. 1). The geniculocortical correlations (A $\rightarrow$ C, B $\rightarrow$ C) indi-

cate that the X cells were monosynaptically connected to the simple cell^{3,4}.

We examined a potential role for the simultaneous firing of the two thalamic neurons by dividing their spikes into two categories: those that occurred within 1.0 ms of each other (A&B) and those of either neuron that did not (A* and B*). If the separate influences from thalamic neurons A and B on the cortical cell were added together independently, then the peak in the summed correlogram from the non-simultaneous spikes (Fig. 3; (A* $\rightarrow$ C) + (B* $\rightarrow$ C)) should resemble the peak in the correlogram for the simultaneous spikes (A&B $\rightarrow$ C). Instead, the size of the peak (above the stimulus-dependent portion, red line) was much greater for the simultaneous spikes. As calculated from these correlograms, the efficacy of simultaneous spikes (A&B, efficacy $5.28 \pm 0.23\%$) was larger than the summed efficacy of non-simultaneous spikes (A* and B*, summed efficacy $3.12 \pm 0.30\%$).

We studied thirteen cases where two geniculate cells were connected to a common simple cell. In each case, simultaneous thalamic spikes (A&B) were more effective in driving the cortical neuron than non-simultaneous spikes (A* or B*) from either neuron. Further, in 11 of 13 cases, the efficacies of simultaneous spikes (A&B) were greater than the sum of the efficacies of A* + B* (mean 71%, median 50%; $P < 0.01$, paired student *t*-test). Included in the 11 cases were two pairs of geniculate cells with completely overlapped receptive fields that showed precisely correlated firing. The remaining 9 cases were from geniculate cell pairs in which both cells were connected to a common cortical neuron, but were not themselves correlated. These cases serve as an important control, as discussed below.

Two possible mechanisms that might be responsible for the

TABLE 1 Percentage of positively correlated LGN cell pairs and correlation strengths		
Overlap of RF centres	Per cent positively correlated	Strength of positive correlations*
No overlap	0% (0/73)	—
Partial overlap	XX: 0/43; XY: 0/28; YY: 0/2	1.9% (n = 20)
	17% (28/167)	
Total overlap	XX: 17/101; XY: 7/51; YY: 4/15	10% (n = 12)
	48% (20/42)	
and same sign	XX: 14/18; XY: 6/23; YY: 0/1	13% (n = 9)
	52% (14/27)	
and similar size	XX: 10/12; XY: 4/14 YY: 0/1	23% (n = 5)
	79% (10/13)	
and similar timing	XX: 10/12; YY: 0/1	28% (n = 4)
	100% (6/6)	
	XX: 6/6	XX: 4

Percentage of positively correlated geniculate cells and correlation strengths as a function of receptive field overlap. Cell pairs with nearly identical receptive fields were more frequently correlated and the correlations were very strong. Receptive fields that differed in any respect (position, size, sign or timing) were less likely to be correlated and the correlations were much weaker. Total overlap: when the centres of the two receptive fields were completely superimposed or when one was contained in the other. No overlap: when the centres of the receptive fields shared no common pixels. Partial overlap: all other cases. Similar size: when the difference in the diameter of the two receptive field centres was less than or equal to 0.8° (two pixels). Similar timing: when $|a_0 - a_1| + |b_0 - b_1| + |c_0 - c_1| + |d_0 - d_1|$ was less than three units, where for a given on-centre cell a is the time of the first frame that contains a significant on-response (20% larger than noise), b is the time of the frame with the strongest on-response, c is the time of the first frame with significant off-rebound and d is the time of the frame with the strongest off-rebound. Note that the sample size in columns 'correlation' and 'strength' is different because only cell pairs recorded with different electrodes were used for determining the strength of correlations. Fast correlations were observed with white-noise, drifting gratings and in the absence of visual stimuli.

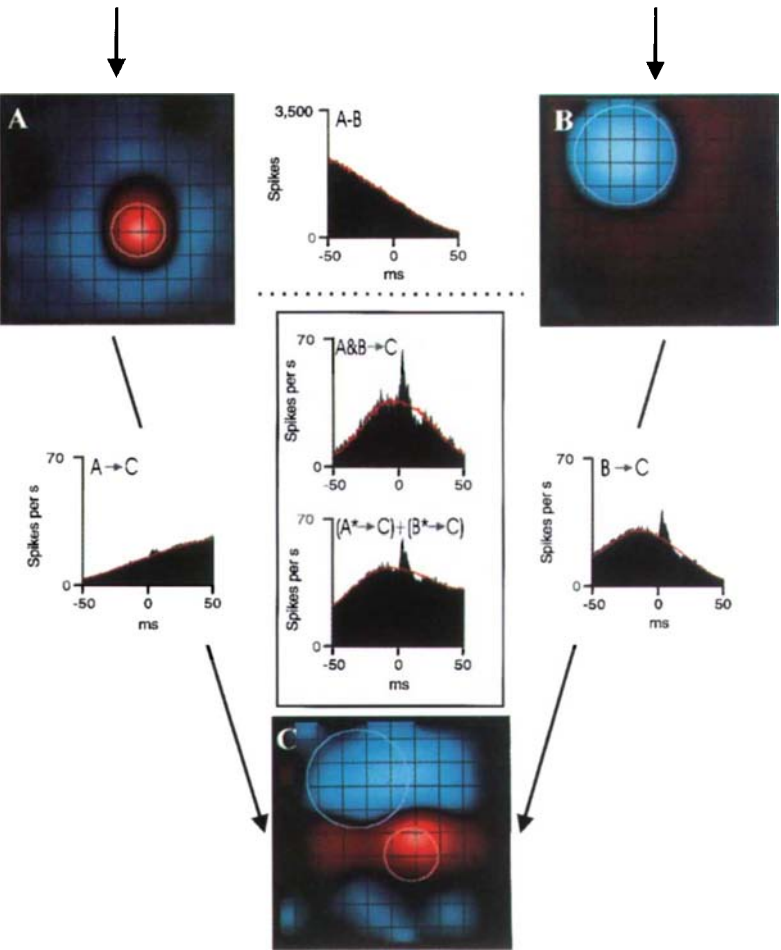
*Strength of correlations calculated only for LGN cells recorded with different electrodes.

increased efficacy of simultaneous spikes are that simultaneous excitatory postsynaptic potentials are more likely to bring the neuron to threshold²⁷, or that they engage some other nonlinear biophysical mechanism in the soma or dendrites (such as calcium action potentials²⁸). Our study cannot distinguish between these possibilities. An alternative explanation is that other strongly correlated geniculate inputs (not recorded) might be responsible for the observed larger efficacy of simultaneous spikes.

Our results indicate, however, that other strongly correlated inputs alone cannot explain the increased efficacy of simultaneous spikes. Non-correlated pairs of geniculate cells also produce simultaneous spikes (A&B), but these events occur infrequently and randomly. There is little chance, therefore, that these random simultaneous spikes are strongly correlated with the spikes of other geniculate neurons. However, in 9 of 11 cases, random simultaneous spikes were more effective in driving a simple cell. One such example is shown in Fig. 4. A pair of non-overlapping X and Y geniculate cells of opposite sign (cells A and B) were recorded from simultaneously, along with an appropriately overlapped simple cell (C). As expected, the geniculate cells were not tightly correlated in their firing. Nevertheless, the efficacy of the simultaneous spikes (A&B, efficacy $5.51 \pm 0.32\%$) was larger than the summed efficacy of the non-simultaneous spikes (A* and B*, summed efficacy $3.60 \pm 0.22\%$). As emphasized by the control shown in Fig. 4, nearly simultaneous spikes do occur by chance. However, the overall frequency of synchronized spikes is increased by the systematic correlations reported here.

It must be noted that whereas synchrony may increase the influence of thalamic neurons on their postsynaptic target, it may lead to overestimates of efficacy in cases where only a single geniculate and cortical cell are studied^{3,4}. Given the size

FIG. 4 Two uncorrelated geniculate cells (*on*-centre X, *off*-centre Y) connected to the same simple cell. The receptive fields have been smoothed by one half pixel. Black squares correspond to one stimulus pixel, or 0.4°. Panels A and B show the receptive fields of an *on*-centre X cell and an *off*-centre Y cell. These two geniculate cells do not show strong, tightly correlated firing (A-B correlation). Therefore, this example serves as a control to demonstrate that the increased efficacy seen in Fig. 3 is not merely due to the existence of a small ensemble of strongly correlated cells. C is the receptive field of a cortical simple cell. The geniculocortical correlations (A → C, B → C) indicate that each geniculate cell is monosynaptically connected to the simple cell^{3,4}. The red lines are the stimulus-dependent correlations (shuffle correlations³⁰). Efficacy of A* = 0.46 ± 0.08 ; efficacy of B* = $3.14 \pm 0.13\%$; efficacy of A* + efficacy of B* = $3.60 \pm 0.22\%$; efficacy of A&B = $5.51 \pm 0.32\%$. Number of spikes collected (A: 166,151; B: 54,114; A*: 162,330; B*: 50,300; A&B: 7,635; C: 51,138).



and frequency of intrageniculate correlations (Table 1), however, this overestimate should not be very large. Fast intrageniculate correlations could also lead to qualitative errors. A positive thalamocortical correlation could be found when a geniculate cell was not in fact connected to a particular cortical neuron but instead was highly correlated with another geniculate cell that was connected to the cortical neuron. Such 'false positives' would not strongly affect previous conclusions about the specificity of geniculocortical connections⁴, because strong intrageniculate correlations are seen only between cells with nearly identical receptive fields. It is thus also unlikely that nonlinear interactions between synchronous inputs (Fig. 3) have a direct effect on receptive field properties, such as orientation selectivity.

To summarize, precisely correlated firing could serve to strengthen the thalamic input to cortical cells^{3,4}. This form of strengthening would act in addition to other factors, including: the size and position of geniculate boutons on the postsynaptic cell⁸, active dendritic conductances²⁹, the high firing rate of geniculate afferents, and the slower correlations within the retina and thalamus¹⁴⁻¹⁹. Moreover, intracortical excitation must also play a role in determining the sensitivity of cortical cells to geniculate input⁹⁻¹¹. On the basis of the findings reported here, the divergence from a single ganglion cell to several cells in the LGN would converge again onto an individual cortical simple cell. This reconvergence would secure the transmission of information²⁶ from retina to visual cortex. □

Methods

Recordings were made from anaesthetized and paralysed cats⁴. Nearby geniculate cells were recorded with a single electrode (42 pairs; 16 pairs correlated), and with different electrodes²⁹ (240 pairs; 32 pairs correlated). Spike isolation was based on cluster analysis of waveforms (Datawave Systems, Broomfield, Colorado) and presence of a refractory period (shape of autocorrelations). To test the significance level of the tight correlations between geniculate cells, cross correlations were bandpass filtered between 75 and 700 Hz. The significance

level was set at a probability of 1.2%, assuming a normal distribution in the baseline amplitudes after filtering. The strength of the correlation was obtained from the unfiltered correlograms⁴.

Received 19 April; accepted 4 September 1996.

- Hubel, D. H. & Wiesel, T. N. *J. Physiol. (Lond.)* **160**, 106–154 (1962).
- Chapman, B., Zahs, K. R. & Stryker, M. P. *J. Neurosci.* **11**, 1347–1358 (1991).
- Tanaka, K. *J. Neurophysiol.* **49**, 1303–1318 (1983).
- Reid, R. C. & Alonso, J. M. *Nature* **378**, 281–284 (1995).
- Ferster, D., Chung, S. & Wheat, H. *Nature* **380**, 249–252 (1996).
- LeVay, S. & Gilbert, C. D. *Brain Res.* **113**, 1–19 (1976).
- Peters, A. & Payne, B. R. *Cerebral Cortex* **3**, 69–78 (1993).
- Ahmed, B., Anderson, J. C., Douglas, R. J., Martin, K. A. C. & Nelson, J. C. *J. Comp. Neurol.* **341**, 39–49 (1994).
- Douglas, R. J., Koch, C., Mahowald, M., Martin, K. A. & Suarez, H. H. *Science* **269**, 981–985 (1995).
- Somers, D. C., Nelson, S. B. & Sur, M. *J. Neurosci.* **15**, 5448–5465 (1995).
- Stratford, K. J., Tarczy-Hornoch, K., Martin, K. A. C., Bannister, N. J. & Jack, J. J. B. *Nature* **382**, 258–261 (1996).
- Hamos, J. E., Van Horn, S. C., Raczkowski, D. & Sherman, S. M. *J. Comp. Neurol.* **259**, 165–192 (1987).
- Cleland, B. G., Dubin, M. W. & Levick, W. R. *Nature New Biol.* **231**, 191–192 (1971).
- Stevens, J. K. & Gerstein, G. L. *J. Neurophysiol.* **39**, 239–256 (1976).
- Sillito, A. M., Jones, H. E., Gerstein, G. L. & West, D. C. *Nature* **369**, 479–482 (1994).
- Neuenschwander, S. & Singer, W. *Nature* **379**, 728–732 (1996).
- Rodieck, R. W. *J. Neurophysiol.* **30**, 1043–1071 (1967).
- Mastronarde, D. N. *Trends Neurosci.* **12**, 75–80 (1989).
- Meister, M., Lagnado, L. & Baylor, D. A. *Science* **270**, 1207–1210 (1995).
- Toyama, K., Kimura, M. & Tanaka, K. *J. Neurophysiol.* **46**, 191–201 (1981).
- Ts'o, D. Y., Gilbert, C. D. & Wiesel, T. N. *J. Neurosci.* **6**, 1160–1170 (1986).
- Singer, W. & Gray, C. M. *Annu. Rev. Neurosci.* **18**, 555–586 (1995).
- Jones, J. P. & Palmer, L. A. *J. Neurophysiol.* **58**, 1233–1258 (1987).
- Sutter, E. *Adv. Meth. Physiol. Systems Model* **1**, 303–315 (Univ. Southern California, 1987).
- Meister, M., Wong, R. O. L., Baylor, D. A. & Shatz, C. J. *Science* **252**, 939–943 (1991).
- Dan, Y., Alonso, J. M., Usrey, W. M. & Reid, R. C. *Soc. Neurosci. Abstr.* **22**, 1703 (1996).
- Jagadeesh, B., Wheat, H. S. & Ferster, D. *Science* **262**, 1901–1904 (1993).
- Hirsch, J. A., Alonso, J. M. & Reid, R. C. *Nature* **378**, 612–616 (1995).
- Eckhorn, R. & Thomas, U. *J. Neurosci. Methods* **49**, 175–179 (1993).
- Perkel, D. H., Gerstein, G. L. & Moore, G. P. *Biophys. J.* **7**, 419–440 (1967).

ACKNOWLEDGEMENTS. We thank T. N. Wiesel for helpful discussion at all stages of this project. J. Hirsch, M. Sanchez-Vives and K. McGowan made useful comments on the manuscript. Technical assistance was provided by K. McGowan, C. Gallagher and D. Landsberger. This study was supported by the NIH, the Klingenstein Fund, Fulbright/MEC and the Charles H. Revson Foundation.

CORRESPONDENCE and requests for materials should be addressed to J.M.A. (e-mail: jma@rockvax.rockefeller.edu).

Loss of morphine-induced analgesia, reward effect and withdrawal symptoms in mice lacking the μ -opioid-receptor gene

Hans W. D. Matthes*, Rafael Maldonado†, Frédéric Simonin*, Olga Valverde‡, Susan Slowe‡, Ian Kitchen‡, Katia Befort*, Andrée Dierich§, Marianne Le Meur§, Pascal Dollé§, Eleni Tzavara||, Jacques Hanoune||, Bernard P. Roques† & Brigitte L. Kieffer*

* UPR 9050 CNRS, ESBS Université Louis Pasteur, 67400 Illkirch, Strasbourg, France

† Département de Pharmacochimie Moléculaire et Structurale, INSERM U266, URA D1500 CNRS, Université René Descartes, 75270 Paris, France

‡ Receptors and Cellular Regulation Research Group, School of Biological Sciences, University of Surrey, Guildford, Surrey GU2 5XH, UK

§ Institut de Génétique et Biologie Moléculaire et Cellulaire, BP 163, 67404 Illkirch, Strasbourg, France

|| Unité de Recherches INSERM U-99, Hôpital Henri Mondor, 94010 Créteil, France

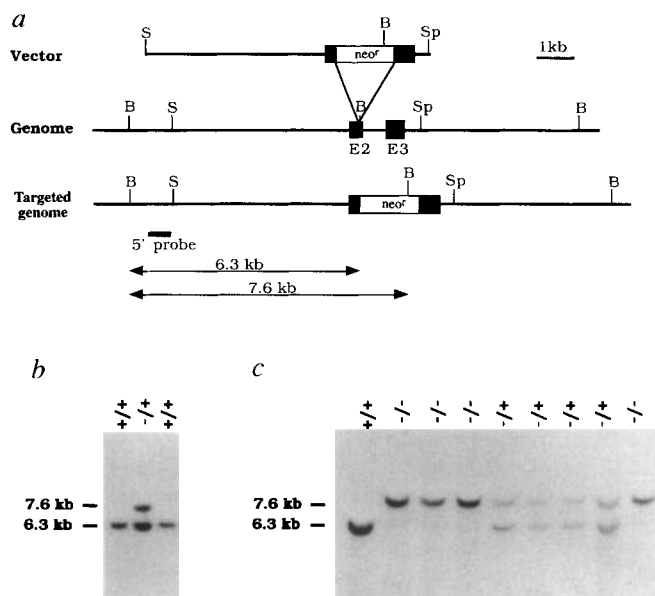
DESPITE tremendous efforts in the search for safe, efficacious and non-addictive opioids for pain treatment, morphine remains the most valuable painkiller in contemporary medicine. Opioids exert

their pharmacological actions through three opioid-receptor classes^{1,2}, μ , δ and κ , whose genes have been cloned³. Genetic approaches are now available to delineate the contribution of each receptor in opioid function *in vivo*. Here we disrupt the μ -opioid-receptor gene in mice by homologous recombination and find that there are no overt behavioural abnormalities or major compensatory changes within the opioid system in these animals. Investigation of the behavioural effects of morphine reveals that a lack of μ receptors abolishes the analgesic effect of morphine, as well as place-preference activity and physical dependence. We observed no behavioural responses related to δ - or κ -receptor activation with morphine, although these receptors are present and bind opioid ligands. We conclude that the μ -opioid-receptor gene product is the molecular target of morphine *in vivo* and that it is a mandatory component of the opioid system for morphine action.

The μ -opioid-receptor (MOR) gene was inactivated in P1 embryonic stem (ES) cells from the 129/Sv mouse line by insertion of a neomycin-resistance (*neo*^r) cassette in the gene coding region (Fig. 1a). Gene targeting was identified by Southern blotting using a 5' probe (Fig. 1b), giving 7 positives out of 90 neomycin-resistant clones analysed. Further Southern analysis using a 3' probe and the *neo*^r probe confirmed accurate integration of a single copy of the disrupted MOR gene fragment (not shown). One positive ES clone was used to establish mutant mice (Fig. 1c).

Mice genotyping showed that heterozygous offspring followed the expected mendelian frequency, with 28.6% wild-type, 46.6% heterozygous and 24.8% homozygous animals ($n = 178$), indicating that there was no *in utero* or postnatal mortality of mutant MOR mice. No obvious morphological abnormalities could be detected in homozygous mutant mice. MOR-deficient mice did not differ from their littermates in health or growth. Homozygous mice were fertile and bred normally, with no impairment of maternal behaviour.

FIG. 1 Disruption of the μ -opioid receptor gene. *a*, Knockout strategy. Restriction maps are shown for the targeting construct (top), wild-type (middle) and recombinant (bottom) gene locus. E2 and E3, second and third coding exons; *neo'*, neomycin-resistance gene; restriction sites: B, *Bam*HI; S, *Sal*I; Sp, *Spe*I. Arrows show *Bam*HI restriction fragments from wild-type (top) and recombinant (bottom) genomes detected by the 5' probe. *b* and *c*, Southern blot analysis of *Bam*HI-digested genomic DNA from electroporated ES cells (*b*) and mouse-tail DNA (*c*), using the 5' probe. Genotypes: +/+, wild type; +/-, heterozygous; -/-, homozygous.



Saturation binding with the μ -selective ligand [3 H]DAMGO showed a 60% reduction of μ receptor in whole brains of heterozygous animals and total loss in homozygous mutants (Fig. 2*a*). Autoradiographic mapping of μ receptors confirmed these results throughout the brain (Fig. 2*b*), demonstrating that our gene targeting strategy led to a complete loss of μ -opioid receptors.

The finding that no DAMGO binding sites are detected in -/- mutant mice brains suggests that both putative μ_1 and μ_2 (ref. 4) receptor sites would arise from the MOR gene.

To investigate the possibility that the absence of the μ receptor may alter expression of other opioid receptors, we quantified the number of δ - and κ -binding sites on whole-brain membranes.

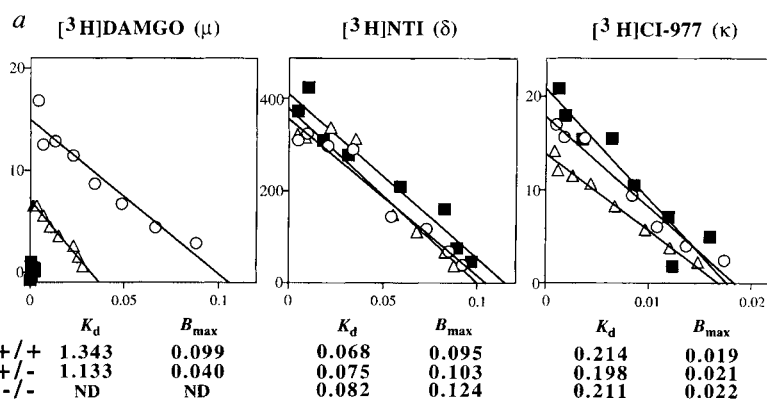
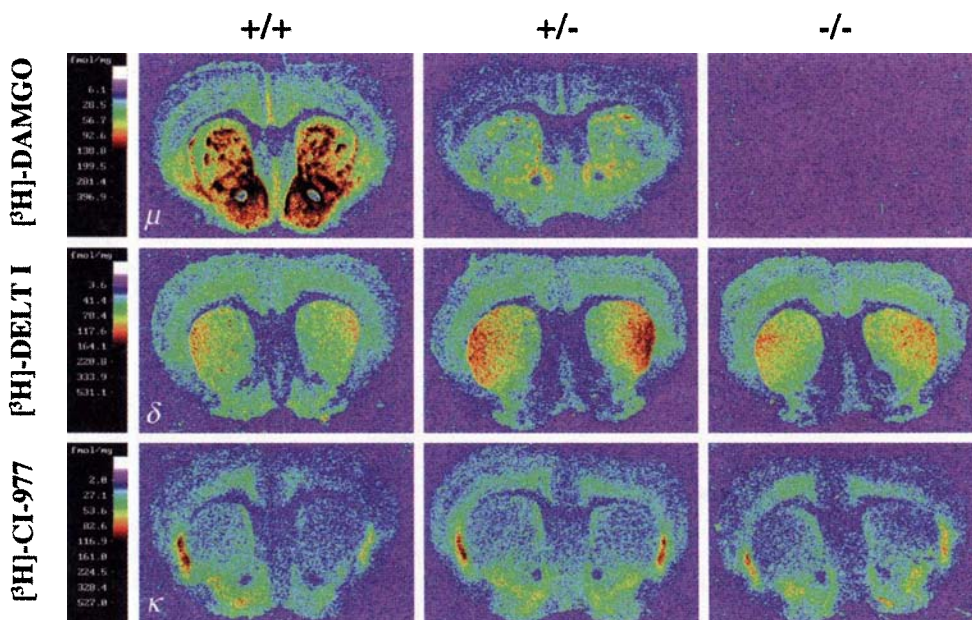


FIG. 2 Expression of μ -, δ - and κ -opioid receptor sites in +/+ (circles), +/- (triangles) and -/- (squares) mice. *a*, Scatchard plots of ligand binding on brain membranes. B_{max} (pmol per mg protein) and K_d (nM) values were obtained from at least three different experiments. ND, not detectable. *b*, Autoradiographic mapping. Computer-enhanced colour autoradiograms of coronal sections show μ , δ and κ binding at the level of caudate putamen (fmol per mg tissue). Ligands: CI-977, [5R-(5 α , 7 α , 8 β)-N-methyl-N-[7-(1-pyrrolidinyl)-1-oxaspiro[4,5]dec-8-yl]benzo[b]furan-4-acetamide; DAMGO, D-Ala²MePhe⁴Gly⁵enkephalin; DELT I, D-Ala²-deltorphin I; NTI, naltrindole.



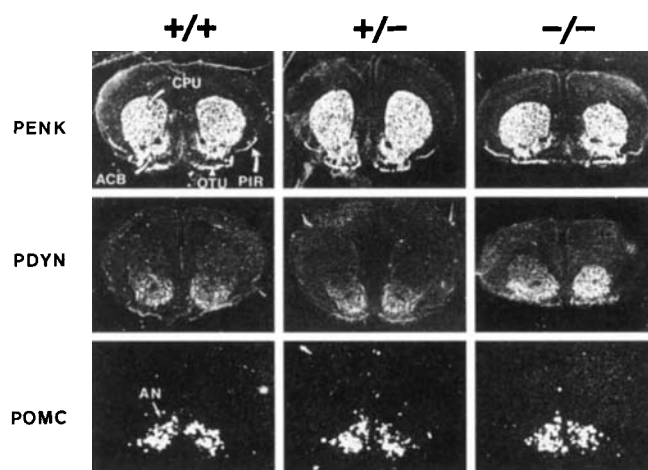


FIG. 3 *In situ* hybridization analysis of endogenous opioid peptide genes in $+/+$, $+/-$ and $-/-$ brains. Coronal sections at the level of the caudate putamen (PENK and PDYN; magnification, $\times 15$) and at the level of the hypothalamus (POMC; magnification, $\times 40$) are shown under dark-field illumination, with the signal grain appearing in white. Antisense probes: PENK, proenkephalin; PDYN, prodynorphin; POMC, proopiomelanocortin; ACB, nucleus accumbens; CPU, caudate putamen; OTU, olfactory tubercle; PIR, piriform cortex; AN, arcuate nucleus.

Scatchard analysis using specific δ -(NTI) and κ -(CI-977) radiolabelled ligands showed similar plots for $+/+$, $+/-$ and $-/-$ mouse brains (Fig. 2a), indicating that the total number of δ and κ receptors is not modified in the brain of μ -receptor-deficient animals. As their discrete distribution may nevertheless be modified, we did a complete autoradiographic mapping using [3 H]DELT I(δ) and [3 H]CI-977(κ) as selective ligands. For both heterozygous and homozygous animals, we found δ - and κ -receptor sites in all the regions where binding was detected in wild-type mice (Fig. 2b). Thus, both the number and distribution of δ and κ sites appears to be unaffected by a lack of μ receptors.

We examined the effect of the absence of μ -opioid receptor on the transcription of endogenous opioid peptide genes⁵ using *in situ* hybridization analysis (Fig. 3). There was no modification in the expression pattern of proenkephalin (PENK) and prodynorphin (PDYN) messenger RNAs, particularly in fore- and midbrain regions where these transcripts are most abundant. Proopiomelanocortin (POMC) expression was limited to arcuate neurons in both wild-type and mutant brains, with similar numbers and distribution of cells containing the transcript. Comparable POMC expression was also observed in intermediate and anterior lobes of the pituitary gland in the two groups of animals (not shown). These results suggest that μ -receptor expression does not exert any regulatory control on the transcription of genes encoding endogenous ligands.

We investigated the consequence of the absence of μ receptor on animal spontaneous behaviour (not shown). A tendency to hypolocomotion was seen in homozygous mutant mice when they were exposed to locomotor activity boxes during the first two habituation sessions. This response was significant in the third session (22% decrease, $P < 0.01$). When recording locomotor activity during day and night periods, mutant and wild-type mice showed a similar increase in locomotor activity during the night, suggesting that the circadian rhythm was not modified. These results show that lack of μ receptors has little influence on locomotor activity, in agreement with the weak or absent effect obtained after administration of μ -opioid antagonists in mouse⁶.

We examined several pharmacological responses of homozygous mutant mice to morphine. Antinociceptive effects of morphine in the tail-immersion and hot-plate tests are shown in Fig. 4. Nociceptive thresholds of mutant and wild-type mice were similar

in both tests, suggesting that endogenous activation of μ receptors is not essential to the control of thermal nociceptive perception, at least for sharp stimuli. This does not exclude a possible change in the perception of other nociceptive stimuli, because administration of opioid antagonists in rodents decreases pain perception in some, but not all, nociceptive tests⁷. Morphine administration induced strong dose-dependent analgesia in wild-type mice in the two tests, but had no effect in mutant mice. In the case of the hot-plate test, this is not surprising as the response is mainly controlled by μ -mediated supraspinal mechanisms⁸. However, lack of morphine response in the tail-flick test was unexpected because the latter primarily involves spinal mechanisms with participation of μ , δ and κ receptors⁹ and, at the doses used, morphine should have some activity at δ receptors. Therefore the involvement of δ sites in spinal-induced analgesia is called into question. Also, administration of the selective δ agonist BUBU¹⁰ did not induce any antinociception in mutant mice in the tail-flick test, whereas the analgesic response was significant in wild-type animals (not shown).

The reinforcing properties of morphine^{11,12} were investigated in the same group of animals using the place-conditioning paradigm¹³. Morphine induced place preference in wild-type mice, whereas this conditioned behaviour was not observed in mutant mice. Mutant mice spent the same time in the morphine-designed compartment during the preconditioning and the testing phase (Fig. 5), presumably because of loss of the rewarding properties of morphine. Therefore, as for spinal analgesia, the rewarding effect of morphine, believed to be mediated by several opioid receptors¹⁴, is abolished in μ -deficient animals.

We induced a severe degree of dependence by giving increasing doses of morphine. Wild-type, but not mutant, animals showed classical signs of chronically opioid-treated mice, such as Straub-tail reflex and increased locomotor activity (not shown). Withdrawal symptoms were evaluated after naloxone administration.

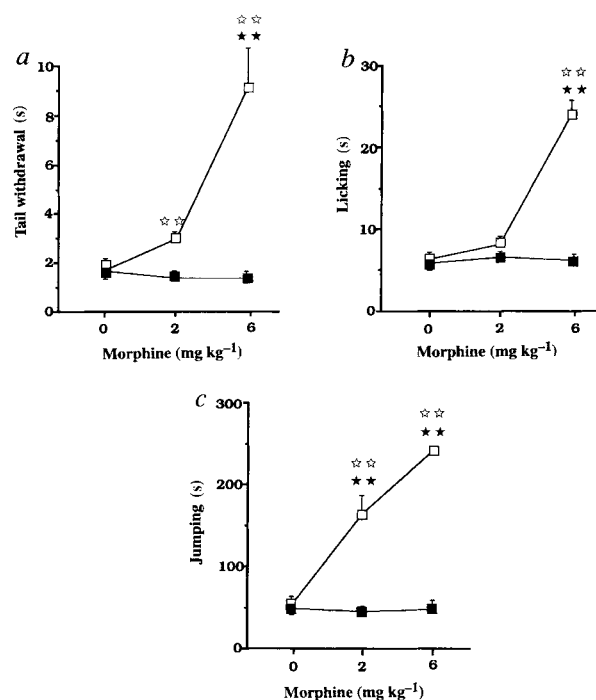


FIG. 4 Antinociceptive responses to morphine administration in mutant mice lacking the μ -opioid receptor ($-/-$; filled squares) and their wild-type littermates ($+/+$; open squares). a, Tail immersion test; b, c, hot-plate test. Values on y-axis represent the latencies in seconds of the different nociceptive thresholds expressed as means $\pm$ s.e.m. Black stars, comparison with saline-treated animals of the same genotype; white stars, comparison between wild-type and mutant groups receiving the same treatment (two-tailed Dunnett's test). Two stars, $P < 0.01$.

All classical signs were seen in morphine-treated wild-type mice. In mutant animals, as in saline-injected controls, naloxone failed to induce any withdrawal (Fig. 6a, b). As chronic opioid treatment upregulates responsiveness of the cyclic AMP signal-transduction pathway, we quantified basal- and forskolin-stimulated adenylyl cyclase activity in wild-type and mutant mice after withdrawal (Fig. 6c). In wild-type mice, there was a 30% increase of both activities in striatum, but not in cerebellum devoid of μ receptors (not shown), as previously described¹⁵. Upregulation of cyclase activity in striatum was absent in μ -deficient animals. These results indicate that absence of the μ receptor in homozygous mutant mice prevents development of physical dependence. Although μ -opioid receptors are critical for the development of opioid physical dependence, δ and κ receptors may also be involved^{16–18}. Here we show that, although the high doses of morphine and naloxone used here are presumably able to stimulate all opioid receptors, morphine physical dependence is totally suppressed in the absence of the μ receptor.

In conclusion, disruption of the μ -opioid receptor gene in mice does not markedly alter expression of other components of the opioid system. This does not exclude the possibility that compensatory changes have occurred in other neurotransmitter systems. We have shown that the MOR gene product is the molecular target of morphine *in vivo*. In addition, we demonstrate that δ and κ receptors do not mediate, even partially, any of the major biological actions of morphine in the absence of the μ receptor. This also raises the important issue of cooperativity between opioid receptors¹⁹, which might take place at the molecular level through receptor allosteric interactions or second messenger systems, or occur at a functional level on separate neurons. Further study of μ -opioid receptor-deficient mice will open new perspectives for understanding the respective roles of μ , δ and κ receptors in the biological actions of opioids. □

Methods

Gene targeting. A mouse genomic library derived from the 129/sv strain was screened using a mouse δ -opioid receptor cDNA²⁰ as probe. A μ -opioid-receptor-encoding genomic fragment containing exons 2 and 3 was obtained and a 6.8-kb *Sall*/*SpeI* fragment was excised and subcloned into pBluescript (Stratagene). A 1.6-kb *Bgl*II cassette containing the neomycin-resistance gene under the control of the PGK promoter²¹ was inserted into a unique *Bam*HI site present in exon 2, thus eliminating the endogenous *Bam*HI site and disrupting the coding sequence at the level of the second intracellular loop (isoleucine 193). The targeting vector was linearized and electroporated as described²². Neomycin-resistant clones were screened by Southern blot analysis using 5', 3' probes and *neo* probes. A positive clone microinjected into C57BL/6 blastocysts²² gave rise to chimaeric offspring, which in turn were mated with C57BL/6 mice. Agouti-coat pups were genotyped by Southern blot analysis of tail DNA for germline transmission.

Binding on brain homogenates. Binding was performed as described²⁰ using 100–150 μ g total brain membrane proteins²³ with [³H]DAMGO (Amersham), [³H]NTI (gift from A. Borsodi) or [³H]CI-977 (Amersham). Experiments were done in triplicate using at least two distinct membrane preparations, each made from three brains. Binding data were analysed using the EBDA-LIGAND program (Biosoft).

Autoradiographic mapping. Mice were decapitated and their brains processed for binding and autoradiography as described²⁴. Adjacent sections (at 300 μ m intervals) were cut for determination of total and nonspecific binding for μ , δ and κ sites labelled with [³H]DAMGO (4 nM), [³H]DELT I (7 nM) (Zeneca, custom synthesis) and [³H]CI-977 (2.5 nM), respectively. Slides were exposed to ³H-Hyperfilm (Amersham) for a period of three (μ and δ) or four (κ) weeks. All slides from +/+, +/- and -/- brains were laid down against the same film, developed and analysed by video-based computerized densitometry (MCID, Imaging Research, Canada). Specific binding was >85% for μ and δ labelling and >75% for κ labelling in regions with high binding.

In situ hybridization. Cryostat sections (10 μ m) were prepared and hybridized as described²⁵. The PENK probe was synthesized from a 800-bp *Pvu*II-*Xba*I cDNA fragment cloned into pBluescript, the PDYN probe from a 1,700-bp *Pst*I-*Eco*RI fragment subcloned into pSP64, and the POMC probe from a 400-bp *Not*I-*Nco*I fragment cloned into pBluescript. cDNAs were a gift from E. Borrelli.

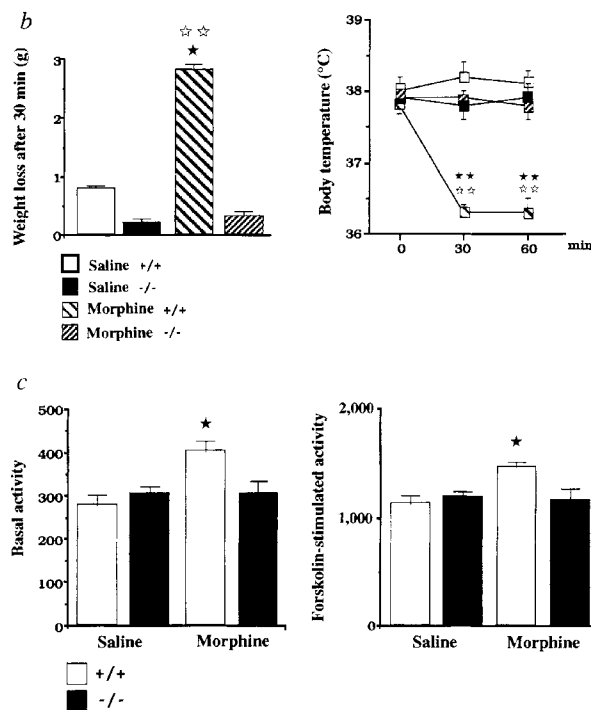
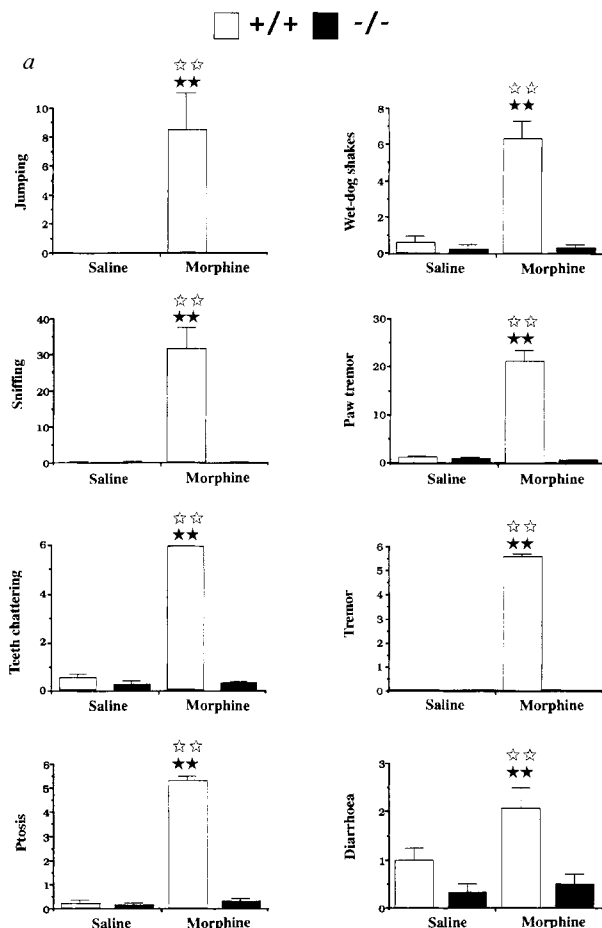


FIG. 6 Naloxone-precipitated morphine-withdrawal syndrome in mutant mice lacking the μ -opioid receptors (-/-) and their wild-type littermates (+/+). a, Somatic signs; b, vegetative signs; c, basal- and forskolin-stimulated adenylyl cyclase activity in striatal homogenates. Results are expressed as means $\pm$ s.e.m.. Black stars, saline-treated animals of the same genotype; white stars, wild-type and mutant groups receiving the same treatment (two-tailed Student's *t* test). One star, $P < 0.05$; two stars, $P < 0.01$.

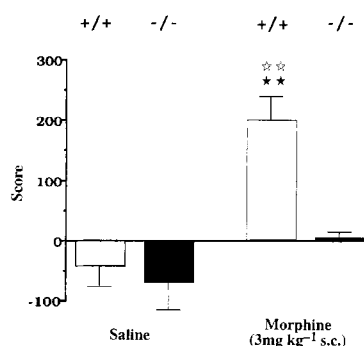


FIG. 5 Conditioned place preference induced by morphine in mutant mice lacking the μ -opioid receptor ($-/-$, black bars), and their wild-type littermates ($+/+$, white bars). Data are expressed in scores calculated as the difference between postconditioning and preconditioning time spent in the compartment associated with the drug. Black stars, comparisons with saline-treated animals of the same genotype; white stars, comparison between wild-type and mutant groups receiving the same treatment (two-tailed Student's *t* test). Two stars, $P < 0.01$.

Analgesia. Tail-immersion (54°C) and hot-plate (52°C) tests²⁶ were performed for 10 and 20 min respectively after saline or morphine injection (2 and 6 mg per kg, injected intraperitoneally). The cut-off was 15 s for the tail-immersion test, and 30 and 180 s respectively for licking and jumping response in the hot-plate test. The number of animals in each group was between 7 and 11. Pharmacological tests and care of the animals were in accordance with standard ethical guidelines (NIH, 1985).

Place preference. The place-preference protocol followed experimental conditions previously reported²⁷, except that the conditioning time (18 min) and the size of compartments ($15\text{ cm} \times 15\text{ cm} \times 15\text{ cm}$) were different. The number of animals per group was between 9 and 14.

Withdrawal. Opioid dependence was induced in mice by repeated intraperitoneal injection of morphine, at an interval of 12 h during 6 days. The morphine dose was progressively increased as follows: day 1, 20 mg kg^{-1} ; day 2, 40 mg kg^{-1} ; day 3, 60 mg kg^{-1} ; day 4, 80 mg kg^{-1} ; day 5, 100 mg kg^{-1} ; day 6, only one injection in the morning, 100 mg kg^{-1} . Control mice were treated with saline under the same conditions. Withdrawal was precipitated only once in each animal by injecting naloxone (1 mg kg^{-1} , subcutaneous) 2 h after the last administration. During the 15 min preceding naloxone injection, mice were checked for normal behaviour. Somatic signs of withdrawal were evaluated immediately after naloxone injection during a period of 30 min, as described²⁸. Colonic temperature was measured 2 min before, 30 min after and 60 min after naloxone injection. The number of animals per group was between 9 and 14.

Adenyl cyclase activity. Mice treated chronically with either saline or morphine followed by a single naloxone injection (see above) were killed 1 h later. Tissues were homogenized, homogenates ($40\text{--}100\text{ }\mu\text{g}$ protein) incubated with [$\gamma\text{-}^{32}\text{P}$]ATP and the amount of [$\gamma\text{-}^{32}\text{P}$]cAMP formed was determined²⁹. Protein was determined using the Bio-Rad assay. Results are expressed in pmol cAMP formed per min per mg protein and represent values of four individual experiments run in triplicate.

Statistical analysis. Behavioural and biochemical data were analysed by two-way ANOVA (mutation and treatment) between subjects. Individual comparisons were made by the two-tailed Student's *t*-test or the two-tailed Dunnett's test (depending on the number of groups) after significant interaction between mutation and treatment in ANOVA.

Received 16 July; accepted 18 September 1996.

1. Brownstein, M. *Proc. Natl Acad. Sci. USA* **90**, 5391–5393 (1993).
2. Goldstein, A. & Naidu, A. *Mol. Pharmacol.* **36**, 265–272 (1989).
3. Kieffer, B. L. *Cell. Mol. Neurobiol.* **15**, 615–635 (1995).
4. Pasternak, G. W. *Clin. Pharmacol.* **16**, 1–18 (1993).
5. Rossier, J. *Nature* **298**, 221–222 (1982).
6. Cowan, A. in *Handbook of Experimental Pharmacology Opioids II* (ed. Herz, A.) Vol. 104, 393–414 (Springer, Berlin, 1993).
7. Roques, B. P., Noble, F., Daugé, V., Fournié-Zaluski, M.-C. & Beaumont, A. *Pharmacol. Rev.* **45**, 87–146 (1993).
8. Baamonde, A., Daugé, V., Gacel, C. & Roques, B. P. *J. Pharmacol. Exp. Ther.* **257**, 767–773 (1991).
9. Dickenson, A. H. *Br. Med. Bull.* **47**, 690–702 (1991).
10. Gacel, G., Zajac, J. M., Delay-Goyet, P., Dauge, V. & Roques, B. P. *J. Med. Chem.* **31**, 374–383 (1988).
11. Di Chiara, G. & North, A. *Trends Pharmacol. Sci.* **13**, 185–193 (1992).
12. Koob, G. F. *Trends Biochem. Sci.* **13**, 177–184 (1992).
13. Katz, R. J. & Gormezano, G. *Pharmacol. Biochem. Behav.* **11**, 231–233 (1979).
14. Shippenberg, T. S., Bais-Kubik, R. & Herz, A. *Brain Res.* **436**, 234–239 (1987).

15. Terwilliger, R. Z., Beitner-Johnson, D., Sevarino, K. A., Crain, S. M. & Nestler, E. J. *Brain Res.* **548**, 100–110 (1991).
16. Maldonado, R., Negus, S. & Koob, G. F. *Neuropharmacology* **31**, 1231–1241 (1992).
17. Wei, E. T. *J. Pharmacol. Exp. Ther.* **216**, 12–18 (1981).
18. Cowan, A., Zhu, X. Z., Mosberg, H. I., Omnaas, J. R. & Porreca, F. J. *Pharmacol. Exp. Ther.* **246**, 950–955 (1988).
19. Traynor, J. R. & Elliott, J. *Trends Pharmacol. Sci.* **14**, 84–85 (1993).
20. Kieffer, B. L., Befort, K., Gavériaux-Ruff, R. C. & Hirth, C. G. *Proc Natl Acad. Sci. USA* **89**, 12048–12052 (1992).
21. McBurney, M. W. et al. *Nucleic Acids Res.* **19**, 5755–5761 (1991).
22. Lufkin, T., Dierich, A., LeMeur, M., Mark, M. & Chambon, P. *Cell* **88**, 1105–1119 (1991).
23. Ilien, B. et al. *Biochem. Pharmacol.* **37**, 3843–3851 (1988).
24. Kitchen, I. et al. *J. Pharmacol. Exp. Ther.* **275**, 1595–1607 (1995).
25. Décimo, D., Labouesse, G. & Dollé, P. in *Gene probes 2: A Practical Approach* (eds Hames, B. D. & Higgins, S.) 183–210 (Oxford Univ. Press, 1995).
26. Schmidt, C. et al. *Eur. J. Pharmacol.* **192**, 253–262 (1991).
27. Valverde, O., Fournier-Zaluski, M.-C., Roques, B. P. & Maldonado, R. *Psychopharmacology* **123**, 119–126 (1996).
28. Maldonado, R. et al. *Science* **273**, 657–660 (1996).
29. Hanoune, J., Stengel, D., Lacombe, M. L., Fiedmann, G. & Coudrier, E. J. *Biol. Chem.* **252**, 2039–2045 (1977).

ACKNOWLEDGEMENTS. We thank P. Chambon and J. C. Stoclet for their support; C. Gavériaux-Ruff, J. Chluba de Tapia and F. Saudou for discussions; D. Metzger for the neo cassette and E. Borrelli for PENK, PDYN and POMC probes; B. Schuhbauer, C. Hindelang, S. Heyberger, and the ES cell culture and microinjection services of IGBMC; J. F. Poinier for animal care; F. Ruffenach for oligonucleotide synthesis; S. Vicaire for DNA sequencing; and B. Boulay and J. M. Lafontaine for photography. This work was funded by the Centre National de la Recherche Scientifique and Association pour la Recherche sur le Cancer. S. Slowe held a Pfizer studentship; B.K. received support from the Ministère de la Recherche et de la Technologie and the Institut Universitaire de France; B.K., I.K. and R.M. received a Biomed-2 support from the European Community. O.V. is recipient of a postdoctoral fellowship from the Association pour la Recherche sur le Cancer.

CORRESPONDENCE and requests for materials should be addressed to B.L.K. (e-mail: briki@esbs.u-strasbg.fr).

Imaging of radiocarbon-labelled tracer molecules in neural tissue using accelerator mass spectrometry

R. E. M. Hedges*, Z. X. Jiang*, C. Bronk Ramsey*, A. Cowey†, J. D. B. Roberts‡ & P. Somogyi‡

* Radiocarbon Accelerator Unit, Research Laboratory for Archaeology, 6 Keble Road, Oxford OX1 3QJ, UK

† Department of Experimental Psychology, South Parks Road, Oxford OX1 3UD, UK

‡ MRC Anatomical Neuropharmacology Unit, Department of Pharmacology, Mansfield Road, Oxford OX1 3TH, UK

AUTORADIOGRAPHY is widely and successfully used to image the distribution of radiolabelled tracer molecules in biological samples. The method is, however, limited in resolution and sensitivity, especially for ^{14}C . Here we describe a new method for imaging ^{14}C -labelled tracers in sections of biological tissue. A highly focused beam of gallium ions bombards the tissue, which is eroded (sputtered) into constituent atoms, molecules and secondary ions. The ^{14}C ions are detected in the secondary beam by the most sensitive method available, namely accelerator mass spectrometry¹. The specimen is scanned pixel by pixel ($1 \times 2\text{ }\mu\text{m}$), generating an image in a manner analogous to scanning electron microscopy. The method can thus be regarded as a specialized form of scanning secondary ion mass spectrometry (SIMS), referred to here as SIAMS (ref. 2). We have used SIAMS to localize the neurotransmitter γ -aminobutyric acid (GABA) in thin sections of cerebral cortex, and show that it can generate ^{14}C images that are much improved on ^{14}C autoradiography. A scan takes 10–20 min and reveals individual axons, neurons and glial cells at high sensitivity. In principle, the resolution could be increased by up to tenfold, and the method could be extended to some other nuclides.

The essential arrangement is shown in Fig. 1. Specimens for the current application were in the form of $0.1\text{--}0.5\text{-}\mu\text{m}$ -thick sections

cut from epoxy-embedded tissue, which were scanned by a 24 KeV Ga^+ beam of 1 nA, with a spot size of $\sim 1 \mu\text{m}$ diameter. The scanning area is divided into 100×100 pixels, with a dwell time of 30 ms per pixel, so a complete scan takes about 5 min and sputters $\sim 10\%$ of the material in the field of view. A negatively charged secondary beam resulting from the sputter bombardment is extracted and injected without further mass analysis into the accelerator mass spectrometry (AMS) system. ^{14}C ions in the beam are individually detected with an efficiency of $\sim 25\%$ (with respect to the secondary C^- beam). The AMS system has been optimized for detection of ^{14}C from archaeological samples ($^{14}\text{C}/^{12}\text{C} \sim 10^{-12}$ – 10^{-15}), for which it is normally dedicated³. In this application^{2,4}, we directly inject the secondary beam without prior selection of mass 14, which allows ^{14}C to be unambiguously detected for $^{14}\text{C}/^{12}\text{C} > 10^{-11}$. This approach has the advantage of making the detection efficiency insensitive to changes in the specimen voltage as a result of charge build-up on the partially insulating surface of the specimen caused by the primary ion beam. A typical count rate for detected ^{14}C is 50 Hz for a $^{14}\text{C}/^{12}\text{C}$ ratio of 10^{-6} . The same system permits imaging in ^{13}C or in ^{12}C , in which case the field is imaged simultaneously, whereas imaging in real time (TV frame rates) is possible using the total negative ion beam. In principle, other ion beams may be used for imaging (such as $^{12}\text{C}^{15}\text{N}^-$), but we have not explored these possibilities.

The critical parameters for performance are the spatial resolution of the primary beam, and the ^{14}C signal obtained from a minimal level of radiocarbon label. The primary beam is provided by a VG MIG102 gun, with a limiting resolution of $0.2 \mu\text{m}$ diameter; ion extraction optics and mechanical design limit present operation to $\sim 1 \mu\text{m}$, but submicrometre resolution is feasible. Note, however, that the minimum adequate ^{14}C concentration is inversely proportional to voxel size and hence to the cube of the beam diameter. The minimum adequate ^{14}C concentration at $1 \mu\text{m}$ is determined by the ratio of C^- ions generated by sputtering to carbon atoms sputtered. This depends on ion source operational conditions, but a ratio of ~ 0.01 can be reliably achieved. Thus, taking into account the 25% detection efficiency, for every 400 ^{14}C atoms sputtered, one is detected. It is necessary to pretreat the sample with evaporated caesium, decreasing its work function, in order to achieve this. In practice, the minimum adequate ^{14}C concentration depends on the ^{14}C contrast and scale of the material being imaged, but we generate images with ~ 20 counts per pixel from locally averaged values of $^{14}\text{C}/^{12}\text{C} \sim 10^{-6}$, with the resolution demonstrated here. The AMS system is designed to measure isotope ratios with minimal systematic error while isotopic discrimination during sputtering from epoxy-impregnated tissue is quite constant. This method therefore enables the ^{14}C -label concentration (relative to bulk carbon) to be quantified with negligible systematic error. (To quantify the

absolute abundance, account must be taken of the local epoxy concentration and the tissue carbon content.) Our specialized form of scanning secondary ion mass spectrometry (SIAMS) could be applied to other nuclides, provided they can be detected with comparably high efficiency. Candidate nuclides include ^3H and ^{36}Cl , and provided label concentrations can be sufficiently increased over natural background abundance, ^{17}O , ^{36}S and ^{15}N could also be usefully imaged by SIAMS.

The cerebral cortex in the rat brain was chosen to test the potential of the method, because it is a very heterogeneous system consisting of many functionally and neurochemically distinct cellular elements, and requires high-resolution methods to dissect its components. We introduced into the brain a ^{14}C -labelled amino-acid neurotransmitter, GABA (4-amino-*n*-[^{14}C]butyric acid), which is selectively taken up by those neurons that normally release it as their transmitter and so also have selective GABA re-uptake and transport mechanisms^{5,6}, exploited previously in the neurochemical characterization of cells by autoradiography⁷. Glial cells also take up GABA⁸, contributing to the elimination of neurotransmitter from the extracellular space. The brain tissue was fixed with a mixture of aldehydes that covalently bound the accumulated transmitter to proteins and thus prevented its loss during tissue processing. The use of thin sections (50–500 nm; prepared by epoxy embedding) enabled individual cells to be characterized at high resolution by immunohistochemistry and electron microscopy, as well as by previously developed criteria for cellular categorization. The experimental design allowed the neurochemical characterization of single neurons and glial cells as well as single axons of 0.3 – $0.6 \mu\text{m}$ diameter in the cerebral cortex.

Figure 2 shows serial sections of the rat cortex imaged for ^{14}C for the distribution of GABA, using antibodies or toluidine-blue stain. The SIAMS images have a pixel size of $2 \times 1 \mu\text{m}$, somewhat larger than the primary beam. The (linear) $^{14}\text{C}/^{12}\text{C}$ ratio in each pixel is colour coded (from blue to red). Where no ^{14}C is detected, the ^{12}C intensity image enables epoxy-plus-tissue areas to be distinguished from epoxy-only areas at low contrast, and epoxy-filled ^{14}C -deficient blood vessels can be recognized and used as landmarks. SIAMS imaging reveals the presence of high concentrations of [^{14}C]GABA in small scattered patches that in consecutive toluidine-blue-stained sections could be identified as neurons (Fig. 2c) or, close to the injection sites, also as glial cells (Fig. 3). Smaller labelled patches revealed in SIAMS were difficult to identify in the toluidine-blue-stained sections used for light microscopy. But as up to two scans of the quality of Fig. 2 were also achieved on sections of $\sim 100 \text{ nm}$ thick, we used transmission electron microscopy on sections consecutive to those scanned in SIAMS to determine whether small profiles were glial cells (Fig. 3), groups of [^{14}C]GABA-containing nerve term-

FIG. 1 The SIAMS system. Secondary ions generated from the sample on the stage (left) are injected (zero field in injection magnet) into the 2 MV tandem accelerator, and $^{14}\text{C}^{3+}$ is detected after mass spectrometry at the analyser magnet and velocity filter. Other ion beams are measured at beam waists F1, F3, F12, F13, by Faraday cups or by particle detectors, as appropriate.

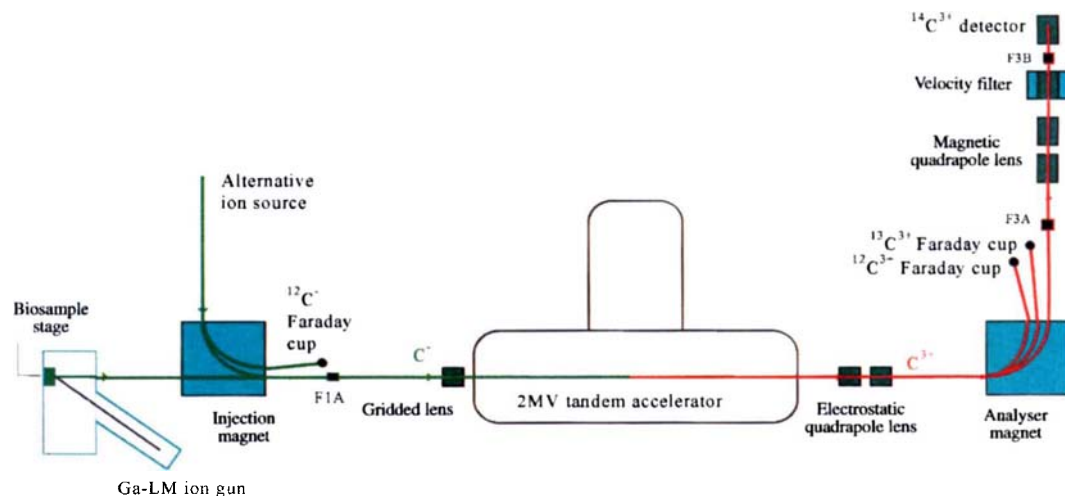
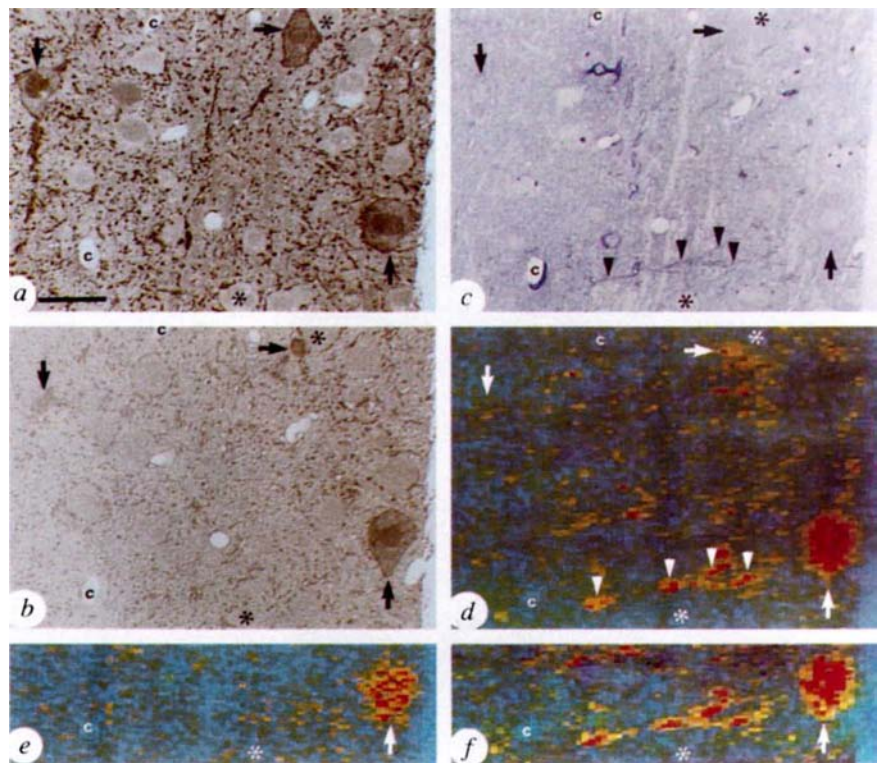


FIG. 2 Identification of [^{14}C]GABA-accumulating neurons in layer 4 of the visual cortex of rat using SIAMS. Serial sections (0.5 μm thick) of the same cells were immunoreacted to reveal GABA using antibodies (a and b, sections 1 and 8 respectively), stained with toluidine blue (c, section 5) or bombarded with a scanning gallium beam of $\sim 1\text{ }\mu\text{m}$ in diameter, followed by AMS to detect the distribution of ^{14}C (d–f). The images on d and e were obtained from section 6 in the series in two separate scans taken several weeks apart. The image in f was obtained from section 3, and all three images are displayed as a ratio of $^{14}\text{C}/^{12}\text{C}$ in each pixel (red highest, blue lowest). There are three GABA-immunopositive neuronal cell bodies (arrows) in this field, one of which (vertical arrow on right) is apparent in all 5 sections and is heavily labelled by [^{14}C]GABA, demonstrating that it selectively accumulates this neurotransmitter. The sectioned area of the cell body of the two other GABAergic neurons diminishes as the series progresses (b), but the uppermost neuron (horizontal arrow) can be identified as labelled even when small. The neuron to the left (vertical arrow) is clearly present in the toluidine-blue-stained section (c), but appears unlabelled in the next section (d). Most other neurons (for example, asterisks) are immunonegative for GABA, but they are surrounded by small brown GABA-immunopositive puncta corresponding to nerve terminals and axons, which are also scattered throughout the neuropil. The presence of ^{14}C in axons and terminals of the neuropil inevitably gives rise to scattered labelling in the SIAMS images as compared to the empty resin at the edge of the material (blue strip, right). The GABA immunonegative neurons (asterisks) and the lumen of capillaries (c) appear as cold islands in the SIAMS images. Some large myelinated axons (arrowheads) of 0.3–0.6 μm diameter (identified in c) are very heavily labelled (d, e). Because only the cytoplasm and not the myelin sheath of the axon contains GABA, and the immunoreacted sections in a and b are several micrometres from the



SIAMS images, it is not possible to identify the location of the axons in a and b. The SIAMS image of the axons appears larger than that in c, because the pixel size ($2 \times 1\text{ }\mu\text{m}$) falsely spreads the label from its real, more restricted site. The area is about 350 μm lateral to the site of [^{14}C]GABA injection. Scale bar, 20 μm .

inals, or small profiles of neuronal cell bodies. For example, in the area shown in Fig. 3, it was difficult to decide which patches of high ^{14}C content corresponded to neurons or glial cells. Electron microscopic examination revealed that two of the patches represented adjoining astroglial cells in direct membrane apposition and giving a labelled area similar in size to that of neurons. Glial cells were identified on the basis of the small smooth outlined nucleus, the thin rim of cytoplasm and the lack of synaptic junctions on their surface as established from serial sections. In contrast, both labelled and unlabelled neurons received numerous synaptic junctions on their soma. In a pattern identical to a previous autoradiographic demonstration of GABA uptake in cortex^{9,10} only a small proportion of neurons were labelled even in the vicinity of the injection tracks. Using antibodies specifically recognizing fixative-modified GABA, it was found that the neurons containing ^{14}C were also immunopositive for GABA. Immunostaining was not solely the result of the injection of radiolabelled GABA, however, because the frequency of immunolabelled neurons was the same in the injected area as in uninjected parts of the cortex, immunostaining of ^{14}C -containing neurons was only rarely enhanced. It appears that the injected transmitter either does not significantly increase GABA levels and/or in the ~ 0.5 –1 h from the injection to fixation of GABA, the radiolabelled pool replaces some of the endogenous pool. In addition, not all neurons that were immunostained for GABA were found to concentrate ^{14}C (Fig. 2), which is to be expected if the uptake of transmitter occurs through the nerve terminals and GABA is then transported retrogradely to the cell body through the axon. Consequently, neurons only accumulate exogenously applied GABA if a large proportion of their terminals are close to the injection site. Different types of GABAergic neurons have

axonal arborizations oriented differentially in a characteristically inter- and intralaminar manner in the cortex^{9–14}. In addition to cell bodies, in several cases ^{14}C -rich hot-spots could be correlated with individual myelinated axons of 0.3–0.6- μm inner thickness, demonstrating the high-resolution potential of the method. These myelinated axons belong to the largest GABAergic cells with the widest axonal arborizations¹², such as one of the cells shown in Fig. 2 (upward-pointing arrow).

We have shown how uptake of labelled GABA by GABAergic neurons can be demonstrated using a new microscopic method. A measurement can be made within 24 hours of sample preparation, takes about 30 min, and is generally applicable to thin sections of epoxy-impregnated tissues. The resolution is limited by that of the primary ion beam (about 1 μm in this work) and the label concentration, but 0.2–0.1 μm , which we did not attempt here, appears feasible. The label concentration required for the present instrument is in the region of $^{14}\text{C}/^{12}\text{C} = 10^{-7}$ at 1 μm and $^{14}\text{C}/^{12}\text{C} = 10^{-4}$ at 0.1 μm . Although the resolution in the current example is comparable to that of using ^3H -labelled molecules for autoradiography¹⁵, such as long exposure time, the limited detection of weak β -radiation from the depth of the sections, and the many complex issues of quantifying the absolute levels of labelled molecules are avoided by SIAMS.

We believe this novel and fast method of atomic analysis of complex heterogeneous systems such as the cerebral cortex has great potential for dissecting molecular composition and interactions, and can be combined for multiple labelling with many other methods such as immunocytochemistry. A further promising application is the quantitative high-resolution cellular mapping of metabolic activity in the brain that takes advantage of the widely

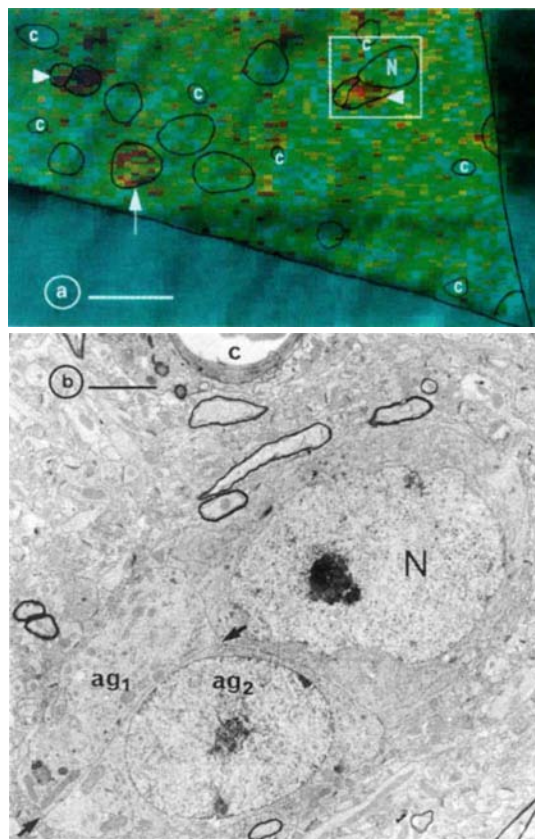


FIG. 3 Identification of [^{14}C]GABA-accumulating cells in layer 4 of the visual cortex of rats using SIAMS (a) correlated with transmission electron microscopy (b) in serial sections (a, 100 nm thick; b, 70 nm thick). a, The corner of an ultrathin section contains three areas of high ^{14}C content of about the same size. Electron microscopic examination revealed that one of them (vertical arrow) corresponds to a non-pyramidal neuron, whereas the other two represent two adjoining ^{14}C -labelled glial cells each (triangles). Black lines outline the edge of the section, capillaries (c), labelled and unlabelled neurons and labelled glial cells as identified in consecutive light and electron microscopic sections. b, The framed area in a is shown in the consecutive electron microscopic section where a capillary, an unlabelled neuron (N) and the two astroglial cells (ag_1 , ag_2) can be identified. The boundary between the glial cells is indicated (arrows). Scale bars: a, 20 μm ; b, 2 μm .

used 2-deoxy[^{14}C]glucose technique¹⁶. SIAMS is applicable to electron microscopic sections of about 80–100 nm thick, which further increases the resolution of the localization as well as the identification of the cellular components. It will be important to increase both the resolution and the sensitivity further for the quantitative localization of ligand binding sites by neurotransmitter receptors and ion channels which can only be localized at present with low resolution. The prospect of using ligands labelled with different isotopes shows how SIAMS imaging can potentially reveal the molecular machinery of chemical communication in the brain. □

Methods

Two adult Wistar rats were deeply anaesthetized with sodium pentobarbital (Sagatal; 45 mg kg⁻¹, i.p., supplemented as required) and mounted in a stereotaxic apparatus. Following a small craniotomy, the occipital cortex was injected unilaterally through glass micropipettes (tip, 30–50 μm) at two sites, perpendicular to the superior sagittal sinus and at 40 degrees to the vertical plane. Radiolabelled GABA (209 mCi mmol⁻¹, Amersham) was dissolved in artificial cerebrospinal fluid at a concentration of 0.95 mM, diluted further by the injection into the brain. Along each track (4–5 mm tangentially, latero-ventral from the pia), 5–8 injections were placed 0.5 mm apart delivering a total of 0.75 μl GABA

(~700 nmol), slowly over 15 min. Animals received an overdose of anaesthetic 50 min after the start of the first injection, followed by transcardial perfusion with a fixative containing 2.5% glutaraldehyde, 0.5% paraformaldehyde in 0.1 M phosphate buffer. The brains were sectioned sagittally (cross sectioning to the tracks), the sections (60- μm -thick) were treated with osmium tetroxide (0.1%) to reduce shrinkage, and embedded in epoxy resin (Durcupan, Fluka)^{10,11}. Strips of cortex (0.2–0.3 mm wide) from pia to white matter at different lateral distances from the injection tracks were sectioned further at 0.5 μm thickness on an ultramicrotome. Serial sections were alternately mounted either on Al stubs for SIAMS, or on glass slides for either the immunocytochemical visualization of GABA¹⁷ using polyclonal antibodies (diluted 1:6,000)¹⁸ and the indirect immunoperoxidase method, or for toluidine-blue-staining to reveal cellular elements in the light microscope. The sections for electron microscopy were picked up on pioloform-coated single-slot grids, and contrasted with uranyl acetate and lead citrate.

Received 1 April; accepted 5 September 1996.

1. Litherland, A. E. *Phil. Trans. R. Soc. Lond. A* **323**, 5–21 (1987).
2. Freeman, S. P. H. T., Bronk Ramsey, C. & Hedges, R. E. M. *Nucl. Instr. Meth. B* **93**, 231–236 (1993).
3. Hedges, R. E. M. *Phil. Trans. R. Soc. Lond. A* **323**, 57–72 (1987).
4. Jiang, Z. X., Bronk Ramsey, C. & Hedges, R. E. M. *Nucl. Instr. Meth. B* (in the press).
5. Martin, D. L. in *GABA in Nervous System Function* (eds Roberts, E., Chase, T. N. & Tower, D. B.) 347–386 (Raven, New York, 1976).
6. Hokfelt, T. & Ljungdahl, A. *Exp. Brain Res.* **14**, 354–362 (1972).
7. Cuenod, M. & Streit, P. in *Methods in Chemical Neuroanatomy* (eds Bjorklund, A. & Hokfelt, T.) 365–397 (Elsevier, Amsterdam, 1983).
8. Hertz, L. *Prog. Neurobiol.* **13**, 277–323 (1979).
9. Somogyi, P., Cowey, A., Halasz, N. & Freund, T. F. *Nature* **294**, 761–763 (1981).
10. Somogyi, P., Kisvarday, Z. F., Freund, T. F. & Cowey, A. *Exp. Brain Res.* **53**, 295–303 (1984).
11. Somogyi, P., Freund, T. F. & Kisvarday, Z. F. *Exp. Brain Res.* **54**, 45–56 (1984).
12. Somogyi, P., Cowey, A., Kisvarday, Z. F., Freund, T. F. & Szentagothai, J. *Proc. Natl Acad. Sci. USA* **80**, 2385–2389 (1983).
13. Somogyi, P., Kisvarday, Z. F., Martin, K. A. C. & Whitteridge, D. *Neuroscience* **10**, 261–294 (1983).
14. Kisvarday, Z. F., Martin, K. A. C., Friedlander, M. J. & Somogyi, P. *J. Comp. Neurol.* **260**, 1–19 (1987).
15. Baker, J. R. J. *Autography: A Comprehensive Overview* (Microscopy Handbooks, Vol. 18) (Oxford University Press, Oxford, 1989).
16. Duncan, G. E. & Stumpf, W. E. in *Quantitative and Qualitative Microscopy; Methods in Neurosciences* Vol. 3 (ed. Conn, P. M.) 50–64 (Academic, New York, 1990).
17. Somogyi, P., Hodgson, A. J., Chubb, I. W., Penke, B. & Erdei, A. *J. Histochem. Cytochem.* **33**, 240–248 (1985).
18. Hodgson, A. J., Penke, B., Erdei, A., Chubb, I. W. & Somogyi, P. *J. Histochem. Cytochem.* **33**, 229–239 (1985).

ACKNOWLEDGEMENTS. We thank the MRC Brain and Behaviour Centre for financial support. We acknowledge valuable discussion with S.P.H.T. Freeman and the contribution of Z. Nusser at the initial stages of test sample preparations.

CORRESPONDENCE and requests for materials should be addressed to R.E.M.H., A.C. or P.S.

The octadecanoid signalling pathway in plants mediates a response to ultraviolet radiation

Antonio Conconi, Michael J. Smerdon, Gregg A. Howe & Clarence A. Ryan*

Institute of Biological Chemistry and Department of Biochemistry and Biophysics, Washington State University, Pullman, Washington 99164, USA

MANY plant genes that respond to environmental and developmental changes are regulated by jasmonic acid, which is derived from linolenic acid via the octadecanoid pathway¹. Linolenic acid is an important fatty-acid constituent of membranes in most plant species and its intracellular levels increase in response to certain signals. Here we report that irradiation of tomato leaves with ultraviolet light induces the expression of several plant defensive genes that are normally activated through the octadecanoid pathway after wounding². The response to ultraviolet light is blocked by an inhibitor of the octadecanoid pathway and it does not occur in a tomato mutant defective in this pathway. The ultraviolet irradiation maximally induces the defence genes at levels where cyclobutane pyrimidine dimer formation, an indicator

of DNA damage, is less than 0.2 dimers per gene. Our evidence indicates that this plant defence response to certain wavelengths of ultraviolet radiation requires the activation of the octadecanoid defence signalling pathway.

The response to ultraviolet (UV) light has been well studied in bacteria³ and yeast⁴ and in animal cell lines at 254 nm wavelength within the UV-C range (below 280 nm), which activates a set of genes that are also induced by UV-A (above 315 nm) and UV-B (280–315 nm) irradiation⁵. UV-C also induces the same photo-products induced in DNA by UV-B⁶ but at higher frequency⁷. To compare the response to ultraviolet irradiation in plants with studies of animals, the induction of two defensive proteinase-inhibitor genes, normally induced in leaves in response to herbivory, was analysed in wounded and unwounded tomato leaves that were exposed to both UV-C and UV-B irradiation.

After exposing the leaves to increasing doses of 254-nm (peak wavelength) ultraviolet irradiation, total RNA was isolated and analysed for proteinase inhibitor I and II messenger RNAs (Fig. 1a). Inhibitor I and II genes were both induced by UV-C in a similar dose-dependent manner, and the highest levels of mRNA were observed after irradiation in the range of 2 kJ m⁻². The levels of the constitutively expressed ubiquitin gene did not change in response to UV irradiation or wounding. Moreover, the synthesis of proteinase inhibitors responded incrementally to increases in UV exposure from 0.5 to 2 kJ m⁻² (Fig. 1b), but at higher doses the effects began to diminish. UV irradiated wounded plants accumulated roughly 50% more inhibitor-I protein than UV irradiated unwounded plants (Fig. 1b), indicating that the UV induction and wound induction were partially additive. Repeated wounding has previously been shown to have an additive effect⁸, therefore UV treatment seems to mimic wounding. Additionally, the time course of the proteinase inhibitor I mRNA induction in response to UV irradiation is similar to that during wound induction (Fig. 1c).

We next explored whether the response of the two proteinase inhibitor genes to UV might be mediated through the octadecanoid pathway, through which several defensive genes (including the inhibitor genes) are regulated in response to wounding^{2,9}. Plants were treated with salicylic acid, a potent inhibitor of the pathway¹⁰, just before UV-C irradiation. Both wound induction and UV induction of the inhibitor I and II proteins were abolished by the presence of salicylic acid (Fig. 2a). A mutant tomato line (JL-5) was recently identified, which carries a mutation in the octadecanoid pathway that blocks the wound-induction of defen-

sive genes¹¹. The UV induction of proteinase inhibitor I mRNA is also blocked in the JL-5 mutant line (Fig. 2b). These findings are consistent and indicate that the UV response of the proteinase inhibitor I and II genes might be mediated by the same intracellular defence-signalling pathway involved in wounding. Several systemic wound-response genes that are activated through the octadecanoid pathway^{2,9} are also UV-C inducible (Table 1). A set of pathogenesis-related proteins (PR1a, PR3a and PR3), which accumulate in response to viral and microbial pathogens and that are not wound inducible¹², are not induced by UV-irradiation in tomato plants. However, a PR1 gene in tobacco plants was recently reported to be induced several hours after UV irradiation^{13,14}. Although the mechanism for this latter response is not known, the UV induction of the pathogenesis-related proteins was attributed to active oxygen species, which have been implicated in the pathogen induction of pathogenesis-related genes in tobacco plants¹⁴ but not in the induction of proteinase-inhibitor genes or other wound-inducible genes.

The UV-induced DNA damage is proposed to be the initial step in the activation of a particular class of genes in animal systems, and has been termed the UV response¹⁵. We looked for an analogous response in plants by measuring UV-C induction of cyclobutane pyrimidine dimer (CPD) formation in the inhibitor I and II genes. The results (Fig. 3) show that an average of about 1 CPD was introduced into the *EcoRI* DNA fragment (~4.2 kilobases) containing the inhibitor I gene (~1.3 kb), when plants were irradiated at 8 kJ m⁻². At UV doses that induce maximal proteinase inhibitor I mRNA and protein accumulation (1–2 kJ m⁻²), roughly 0.2 CPDs were found in the same DNA fragment, and an even lower level is probably introduced into the inhibitor II gene (~30% of the *EcoRI* fragment length). Although the DNA sequences of the proteinase inhibitor I and II genes are unrelated, the two genes are coordinately induced by wounding⁸ and ultraviolet light (Fig. 1). As CPD formation depends on the DNA sequence^{6,16}, we directly compared the UV-induced CPD levels in the inhibitor I DNA fragment (~4.2 kb) with those in the *HindIII* DNA fragment (~5.0 kb) containing the inhibitor II gene (~1 kb). The formation of CPDs in both inhibitor fragments was found to be similar in wounded and unwounded plants (Fig. 3), that is, the maximum expression in both genes occurs at less than ~0.2 CPD per proteinase inhibitor gene. Moreover, CPDs were formed in UV-irradiated JL-5 mutant tomato plants at levels comparable to wild type plants (Fig. 3). It is possible that the yield of other UV photoproducts (for example 6-4-pyrimidine pyrimidones)

FIG. 1 UV dose-response of the proteinase inhibitor I and II in leaves of 14-day-old tomato plants. *a*, Plants were wounded once with a haemostat and UV-irradiated 2.5 h after wounding. Total RNA was isolated from the whole leaves 6.5 h after UV irradiation (9 h after wounding) and analysed for the proteinase inhibitor I (Inh I)²⁸, the proteinase inhibitor II (Inh II)²⁹ or ubiquitin (UBQ) RNA. The lanes correspond to RNA isolated from unwounded (U) or wounded (W) leaves UV-irradiated at a dose of 1, 2, 4 or 8 kJ m⁻². Lanes 0, RNA isolated from control plants. *b*, Unwounded (solid bars) and wounded (hatched bars) tomato plants were UV-irradiated with increasing doses, incubated under standard conditions for 25 h and analysed for proteinase inhibitor I protein content by immunoradial diffusion. Error bars, s.e.m. of six different experiments. *c*, Densitometric quantification of the inhibitor I northern blots. Filled squares, wounding; filled circles, irradiation at 2 kJ m⁻²; open circles, unwounded, non-irradiated control plants.

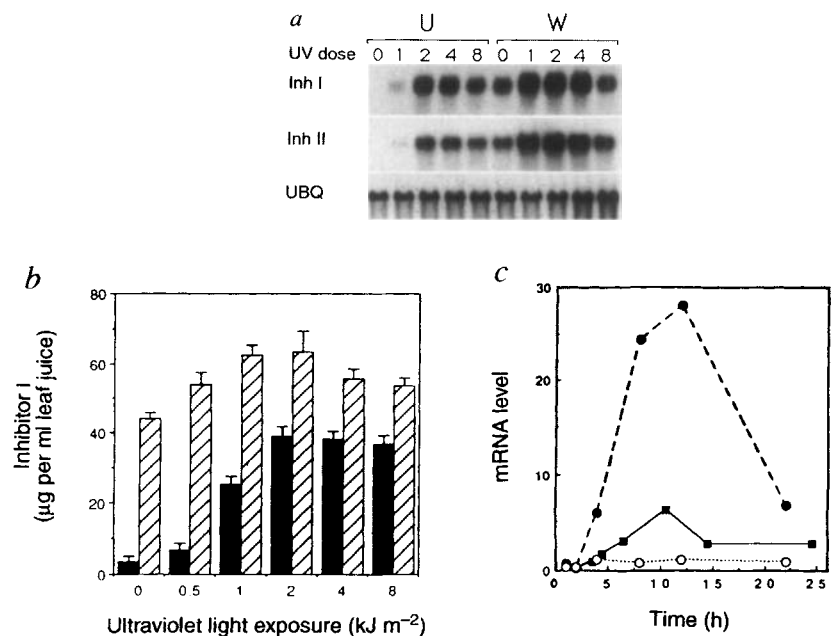
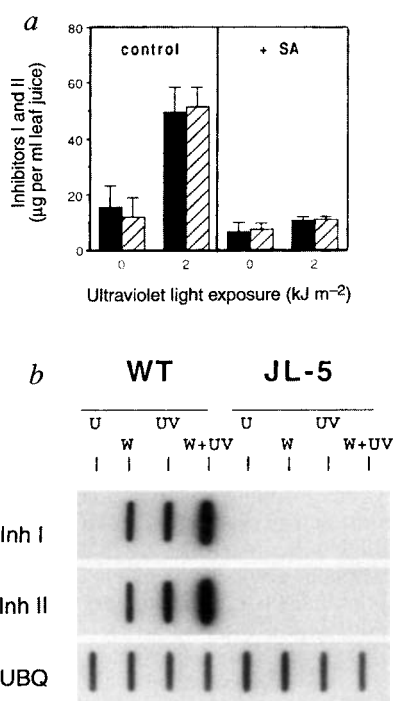


FIG. 2 Inhibition in leaves of 14-day-old tomato plants of wound- and UV-inducible proteinase inhibitor I and II synthesis by salicylic acid and by a mutation in the octadecanoid pathway. *a*, Excised tomato plants were supplied with 1 mM Salicylic acid (+SA) or water (Control) through the cut stems, 1 h before UV irradiation (2 kJ m^{-2}), incubated together with the non-irradiated control plants (0 kJ m^{-2}) for 25 h under standard conditions and analysed for proteinase inhibitor I (solid bars) and II (hatched bars) contents. Error bars, s.e.m. of two different experiments (6 plants each). *b*, RNA was isolated from UV-irradiated (2 kJ m^{-2}) wild-type (WT) and mutant (JL-5) plants. (U) unwounded, (W) wounded, (UV) UV-irradiated and (W + UV) wounded and UV-irradiated leaves. (Inh I) inhibitor I, (Inh II) inhibitor II and (UBQ) ubiquitin mRNAs.



might have been found to correlate differently with the induction response, but these lesions are present at even lower frequencies than CPD in most DNA¹⁷. The relative low concentration of CPD found in the AT-rich (~70%) sequences of the proteinase inhibitor I and II genes, after exposure to very high UV dose, may be explained by the presence in plants of shielding mechanisms (cuticular waxes, trichomes, pigments and flavonoids) that efficiently absorb or scatter ultraviolet light¹⁸.

The proteinase inhibitor genes that are activated in tomato

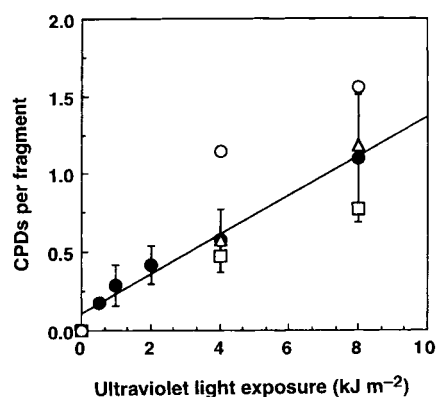


FIG. 3 UV-induced cyclobutane pyrimidine dimer (CPD) formation in genomic DNA fragments containing the proteinase inhibitor I and II genes. *Eco*RI or *Hind*III digested leaf DNA was incubated with T4 endo V nuclease, separated by alkaline agarose gel and probed for the proteinase inhibitor I or II genes. The data represent the average yield of CPDs per 4.2-kb DNA fragment and s.d. for three different experiments obtained for the proteinase inhibitor I DNA fragment in unwounded leaves (filled circles). Average of two different experiments obtained for the unwounded proteinase inhibitor II DNA fragment (triangles), the wound-induced proteinase inhibitor II DNA fragment (squares) and the inhibitor I DNA fragment of JL-5 tomato mutants (open circles).

TABLE 1 UV-C induction of systemic wound-response genes in leaves of tomato plants

Gene product	WT		JL-5	
	Wounded	UV irradiated	Wounded	UV irradiated
Proteinase inhibitor I	+	+	-	-
Proteinase inhibitor II	+	+	-	-
SH proteinase inhibitor	+	+	-	-
Asp proteinase inhibitor	+	+	-	-
Polyphenol oxidase	+	+	-	-
Leucine aminopeptidase	+	+	-	-
Threonine deaminase	+	+	-	-
PR1a	-	-	-	-
PR3a	-	-	-	-
PR3	-	-	-	-
Ubiquitin	(-)	(-)	(-)	(-)

After wounding or UV irradiation of wild-type (WT) and mutant (JL-5) plants, RNA was isolated (7.5 h after UV irradiation). Northern blots were probed with: proteinase inhibitor I and II complementary DNAs²⁹, SH proteinase inhibitor³⁰, Asp proteinase inhibitor³⁰, polyphenol oxidase (from P. Constabel), leucine aminopeptidase (from L. Walling), threonine deaminase³⁰; tobacco acidic PR1a, PR3a and basic PR3 cDNAs (from J. Bol) and *Antirrhinum* ubiquitin (from C. Martin). +, gene induction; -, no gene induction or RNA present at low levels; (-), no gene induction but RNA constitutively present.

leaves by UV-C irradiation are also activated by irradiation with the more ecologically relevant 310 nm UV-B light (Table 2) but are not induced upon UV-A irradiation (data not shown). Although UV-B was less effective than either wounding or UV-C in inducing the synthesis of the proteinase inhibitors I and II, UV-B activation was about three times higher than that observed in leaves of untreated control plants. No activation occurred in leaves of the JL-5 tomato mutant that has an impaired octadecanoid signalling pathway. This indicates that the octadecanoid pathway in tomato plants probably mediates both UV-C and UV-B responses of the proteinase inhibitor I and II genes.

In animal cells, free-radical formation has been shown to occur in response to UV-A and UV-B irradiation¹⁹, potentially leading to membrane perturbation, activation of phospholipases and the release of arachidonic acid²⁰⁻²². Arachidonic acid, the precursor of the prostaglandins in animals, has been shown to be released on UV-C irradiation of cell membranes²³. In plants, which do not contain arachidonic acid or prostaglandins, linolenic acid is released from membranes in response to wounding, and is converted to the prostaglandin analogues phytodienoic acid and jasmonic acid via the octadecanoid pathway². These prostaglandin analogues are powerful inducers of the systemic wound-response

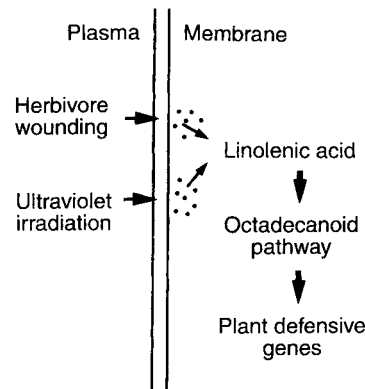


FIG. 4 Proposed model for the signalling of the ultraviolet response in tomato leaves.

TABLE 2 Extent of induction of proteinase inhibitors in leaves

Inducer	Proteinase inhibitor I		Proteinase inhibitor II	
	WT	JL-5	WT	JL-5
Control	4.9 ± 1.2	0	3.2 ± 1.3	0
Wounding	19.7 ± 1.0	0	21.5 ± 1.9	0
UV-C	20.0 ± 1.0	0	20.9 ± 2.1	0
UV-B	13.6 ± 1.0	0	10.9 ± 1.3	0

Comparison of the induction of proteinase inhibitors I and II in leaves of 11-day-old wild-type (WT) and mutant (JL-5) tomato plants by wounding, UV-B and UV-C irradiation. In parallel, under the same light conditions, tomato plants were either wounded, or irradiated with 254 nm UV-C (2 kJ m⁻²) or with 310 nm UV-B (30 to 40 kJ m⁻²). The control and the treated plants were incubated under standard conditions for 25 h and the content of the proteinase inhibitor I and II proteins was analysed by immunoradial diffusion. Mean ± s.e.m. of three different experiments.

genes. To explain the similarities between wounding and the UV response in activating systemic wound-response genes, we suggest that UV irradiation results in the perturbation of plant membranes and/or the activation of lipases to cause the release of linolenic acid, which engages the intracellular octadecanoid signal-transduction pathway to activate the wound inducible genes (Fig. 4). It is also possible that DNA damage might be the initial signal that activates the octadecanoid pathway even though only low levels of DNA damage occurred in tomato leaves upon UV-C irradiation. However, this damage could result in the production of a signal that is released from the nucleus into the cytoplasm which then activates a membrane-bound signalling process, as proposed in animal cell lines²⁴.

A long-term ozone depletion over the mid-latitudes of both the northern and southern hemispheres has resulted in elevated UV exposures over extended periods. Increased UV exposures were reported to alter the competitiveness of some plant species and possibly to alter plant communities and ecosystems over a relatively short evolutionary period²⁵. It is possible that the activation of signalling pathways for defence genes by increased levels of ultraviolet light may contribute to these changes. □

Methods

Tomato plants (*Lycopersicon esculentum* (L.) Mill., cv Castelmart) were grown from seeds under standard conditions (17 h of light (250 µE m⁻² s⁻¹) at 27 °C and 7 h in the dark at 18 °C. We used leaves and cotyledons of 10-day-old plants (two small developed leaves) and irradiation was done with four low-pressure mercury lamps (predominantly 254 nm) at a flux of ~15 W m⁻² (measured with a Spectronics DM-254 N UV meter). The UV-C treated plants did not show any phenotypic evidence of stress, except at UV doses of 8 kJ m⁻², where shiny spots appeared on the leaves. Also, irradiated plants showed a slower growth rate than control plants. The UV-B experiments were done with two UV-B lamps (peak wavelength, 310 nm) and the flux (~35 W m⁻²) was measured with a UVP UV meter.

Wounded and/or UV-irradiated plants were further incubated in the light under standard conditions for different periods of time, and the induction of the proteinase inhibitor I and II genes was measured by northern blots analysis. Alternatively, the proteinase inhibitor I and II protein content was analysed by immunoradial diffusion as previously described⁸. For CPD measurements after UV-C irradiation, unwounded or wounded tomato leaves were collected in the dark and frozen in liquid nitrogen to limit DNA repair by endogenous enzymes. Total leaf DNA was first digested with *EcoRI* or *HindIII* and then incubated with T4 endo V nuclease (which nicks DNA strands at CPD sites) or mock treated²⁶. The average number of CPDs per fragment was determined from the fraction free of CPDs using the Poisson equation ($-\ln(P_0)$) (ref. 27).

Received 15 July; accepted 3 September 1996.

1. Reinbothe, S., Mollenhauer, B. & Reinbothe, C. *Plant Cell* **6**, 1197–1209 (1994).
2. Farmer, E. E. & Ryan, C. A. *Plant Cell* **4**, 129–134 (1992).
3. Walker, G. C. *Microbiol. Rev.* **48**, 60–93 (1984).
4. Jones, J. S. & Prakash, L. *Nucleic Acids Res.* **19**, 893–898 (1991).
5. Cerutti, P. A. & Trump, B. F. *Cancer Cells* **3**, 1–7 (1991).
6. Mitchell, D. L., Jen, J. & Cleaver, J. E. *Nucleic Acids Res.* **20**, 225–229 (1992).
7. Quate, F. E., Sutherland, B. M. & Sutherland, J. C. *Nature* **358**, 576–578 (1992).
8. Graham, J. S., Hall, G., Pearce, G. & Ryan, C. A. *Planta* **169**, 399–405 (1986).
9. Hildmann, T. et al. *Plant Cell* **4**, 1157–1170 (1992).

10. Peña-Cortés, H., Albrecht, T., Prat, S., Weiler, E. W. & Willmitzer, L. *Planta* **191**, 123–128 (1993).
11. Howe, G. A., Lightner, J., Browne, J. & Ryan, C. A. *Plant Cell* (in the press).
12. Bol, J. F., Linthorst, H. J. M. & Cornelissen, B. J. C. *Annu. Rev. Phytopathol.* **28**, 113–138 (1990).
13. Brederode, F. T., Linthorst, H. J. M. & Bol, J. F. *Plant. Mol. Biol.* **17**, 1117–1125 (1991).
14. Green, R. & Fluhr, R. *Plant Cell* **7**, 203–212 (1995).
15. Herrlich, P., Ponta, H. & Rahmsdorf, H. J. *Rev. Physiol. Biochem. Pharmacol.* **119**, 187–223 (1992).
16. Gordon, L. K. & Haseltine, W. A. *Radiation Res.* **89**, 99–112 (1982).
17. Mitchell, D. L., Jen, J. & Cleaver, J. E. *Photochem. Photobiol.* **54**, 741–746 (1991).
18. Stapleton, A. E. *Plant Cell* **4**, 1353–1358 (1992).
19. Black, H. S. *Photochem. Photobiol.* **46**, 213–221 (1987).
20. Hawk, J. L. M. et al. *J. Invest. Dermatol.* **80**, 496–499 (1983).
21. Cohen, D. et al. *Photochem. Photobiol.* **57**, 383–390 (1993).
22. Tyrrell, R. M. *BioEssays* **18**, 139–148 (1996).
23. Sheng, Y. & Birkle, D. L. *Curr. Eye Res.* **14**, 341–347 (1995).
24. Radler-Pohl, A. et al. *EMBO J.* **12**, 1005–1012 (1993).
25. Gold, W. G. & Caldwell, M. M. *Physiol. Plant.* **58**, 435–444 (1983).
26. Murad, A. O., de Cock, J., Brown, D. & Smerdon, M. J. *J. Biol. Chem.* **270**, 3949–3957 (1995).
27. Bohr, V. A. & Okumoto, D. S. *DNA Repair: A Laboratory Manual of Research Procedures* Vol. 3 (eds Friedberg, E. C. & Hanawalt, P. C.) 347–366 (Dekker, New York, 1988).
28. Lee, J. S. et al. *Proc. Natl Acad. Sci. USA* **83**, 7277–7281 (1986).
29. Graham, J. S. et al. *J. Biol. Chem.* **260**, 6561–6564 (1985).
30. Schaller, A., Bergey, D. R. & Ryan, C. A. *The Plant Cell* **7**, 1893–1898 (1995).

ACKNOWLEDGEMENTS. We thank S. Lloyd for providing us with the T4 endo V enzyme, and S. H. Doares and J. W. Smerdon for technical help. This research was supported in part by project 1391, College of Agriculture and Home Economics, Washington State University, by grants from the NSF to C.A.R. and from the NIH to M.J.S.

CORRESPONDENCE and requests for materials should be addressed to C.A.R.

Insertional mutagenesis and rapid cloning of essential genes in zebrafish

Nicholas Gaiano, Adam Amsterdam, Koichi Kawakami, Miguel Allende, Thomas Becker & Nancy Hopkins

Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

LARGE-SCALE chemical mutagenesis screens in zebrafish have led to the isolation of thousands of lethal mutations in genes that are essential for embryonic development^{1,2}. However, the cloning of these mutated genes is difficult at present as it requires positional cloning methods. In *Drosophila*, chemical mutagenesis screens were complemented with P-element insertional mutagenesis which facilitated the cloning of many genes that had been identified by chemical lesions^{3,4}. To facilitate the cloning of vertebrate genes that are important during embryogenesis, we have developed an insertional mutagenesis strategy in zebrafish using a retroviral vector. Here, in a pilot screen of 217 proviral insertions, we obtained three insertional mutants with embryonic lethal phenotypes, and identified two of the disrupted genes. One of these, *no arches*, is essential for normal pharyngeal arch development, and is homologous to the recently characterized *Drosophila* zinc-finger gene, *clipper*, which encodes a novel type of ribonuclease⁵. As it is easy to generate tens to hundreds of thousands of proviral transgenes in zebrafish⁶, it should now be possible to use this screening method to mutate and then rapidly clone a large number of genes affecting vertebrate developmental and cellular processes.

The infection of blastula-stage zebrafish embryos with murine leukaemia virus/vesicular stomatitis virus pseudotyped retroviral vectors^{7,8} results in the germline transmission of proviral insertions^{6,9}. Integration of proviral DNA into the mouse genome can disrupt essential genes¹⁰. In an effort to isolate insertional mutations in zebrafish, we generated and inbred 217 proviral transgenes and screened for mutant phenotypes that were linked to those transgenes. The virus used to generate the insertions

SFG(G) has been described elsewhere⁶. Three proviral transgenes, referred to as 38M, 67D and 80A, were found to be linked to mutant phenotypes as all mutant embryos were transgenic. These three phenotypes are lethal, including a variety of defects in the head, visceral organs and fins, and in all three cases mutant embryos develop oedema by day 5 of development (Fig. 1).

Fragments of the genomic DNA flanking the three potentially mutagenic insertions were isolated, then single-copy sequences were identified and these sequences used as probes for Southern blots to determine the genotype of embryos generated by inbreeding the insertions (Fig. 2). For insertion 38M, we analysed 110 mutants and 272 wild-type siblings; for insertion 67D, 133 mutants and 252 wild-type siblings; and for insertion 80A, 109 mutants and 241 wild-type siblings. In all three cases, every mutant embryo examined was found to be homozygous for the insertion in question, and every phenotypically wild-type sibling was found to be either heterozygous or non-transgenic. With the number of embryos examined in each case, if a mutation and the associated insertion were 1 cM apart, we would have observed 3–4 recombinants. No recombinants were observed, indicating that in all three cases the insertions and mutations are tightly linked.

In an effort to identify genes mutated by these insertions, 2–5 kilobases (kb) of genomic DNA at each insertion site was sequenced, and these sequences were used to search the sequence databases¹¹. Sequences adjacent to the site of insertion 38M were found to be highly homologous to the *Drosophila* gene *clipper* (*clp*) (ref. 5, and see below), as well as to genes of unknown function identified in the *Caenorhabditis elegans* and yeast genome projects. Sequences adjacent to the site of insertion 67D were found to be highly homologous to a human expressed-sequence tag (EST), and to genes of unknown function in *C. elegans* and yeast.

Intron and exon structures for both putative fish genes were predicted within the cloned genomic regions on the basis of the homologous sequences in other organisms (Fig. 3a), and both genes were found to be expressed in zebrafish embryos, using reverse transcriptase with the polymerase chain reaction (RT-PCR) (Fig. 3b). To determine whether the expression of these genes was disrupted by the proviral insertions, northern blot analysis was used (Fig. 3c). RNA isolated from wild-type embryos on day 5 of development was found to contain a 1.9-kb transcript

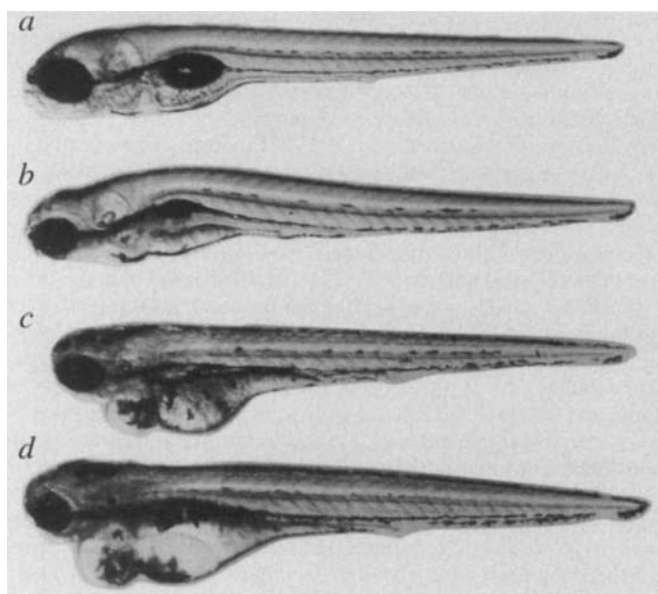


FIG. 1 Comparison of wild type and mutant embryos on day 5 of development. a, Wild-type embryo. b, c, d, Mutant embryos obtained by inbreeding insertions 38M, 67D and 80A, respectively. Inbreeding crosses were screened as described previously⁴.

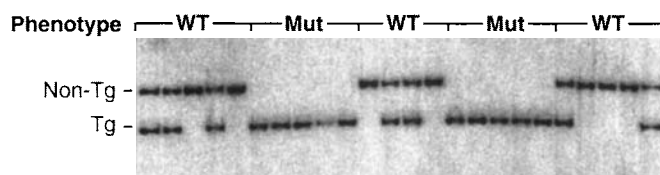


FIG. 2 Linkage analysis between insertion 38M and the mutant phenotype resulting from its inbreeding. Phenotypically wild-type (WT) and mutant (Mut) embryos from a cross of two fish heterozygous for insertion 38M were sorted on day 5, and individual embryos were genotyped by Southern blot. Tg, transgenic.

when probed with a complementary DNA fragment from the gene at the site of insertion 38M, and a 2.3-kb transcript when probed with a cDNA fragment from the gene at the site of insertion 67D (Fig. 3c, lanes 1 and 4, respectively). Insertions 38M and 67D were inbred and RNA was made on day 5 of development from sorted pools of phenotypically wild-type and mutant embryos. For both insertions, the phenotypically wild-type embryos expressed transcripts of the expected size (Fig. 3c, lanes 2 and 5), whereas no transcripts could be detected in the mutant embryos (lanes 3 and 6).

These results demonstrate that the proviral insertions 38M and 67D have either disrupted the expression of the genes into which they have integrated, or in homozygous embryos have resulted in a reduction or deletion of the tissues in which those genes are

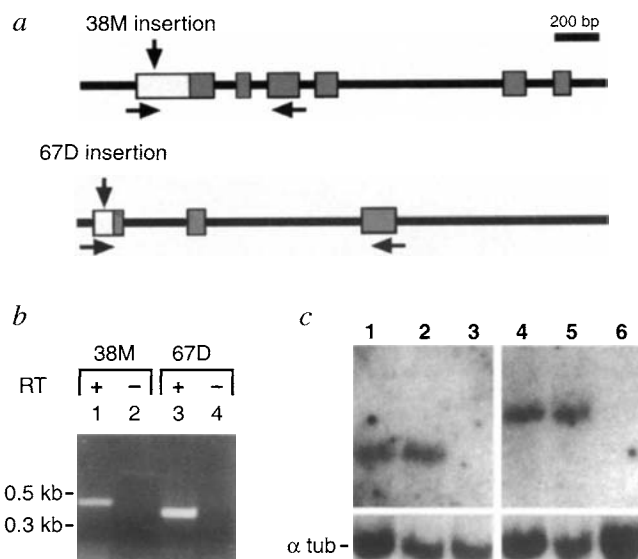
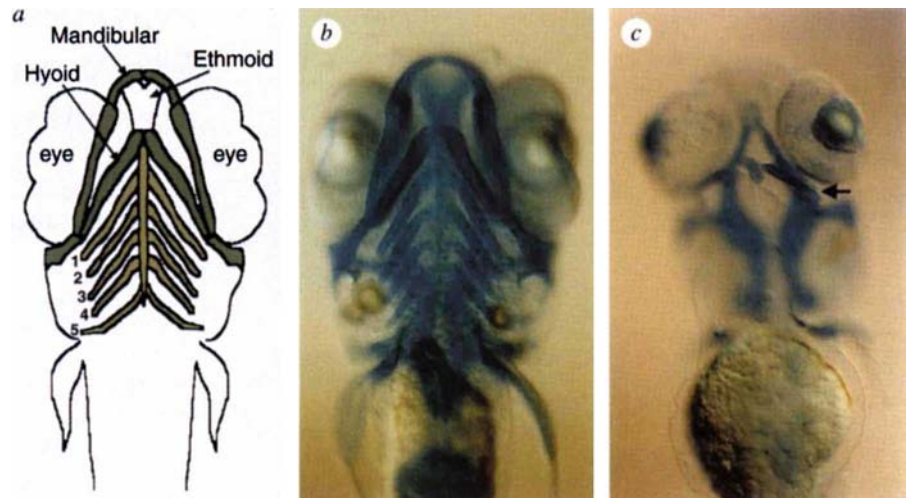


FIG. 3 Provirial insertions 38M and 67D have integrated into and disrupted expressed zebrafish genes. a, The genomic site of insertion 38M encodes amino-acid homology to the *Drosophila* gene *clp* (shaded)⁵. The genomic site of insertion 67D encodes amino-acid homology to a human expressed sequence tag (shaded). In both cases the exons depicted encode only the 5' portion of the fish genes. b, RT-PCR showing that insertions 38M and 67D have both landed in exon sequence. The primers used are shown in a. The identity of these PCR products and the predicted splicing events were confirmed after subcloning and sequencing. In both cases, the amplified sequences extend upstream (white boxes) of the exon sequences identified by homology to other species (shaded). The absence of open reading frames which extend past the homologous regions suggests that the insertions have occurred into the 5'-untranslated region of these genes. c, Northern-blot analysis of embryos on day 5 showing the disruption of zebrafish genes by insertions 38M and 67D. See text for details. An α -tubulin (α tub) probe was used to confirm the presence of RNA in all lanes. RT, reverse transcriptase.

FIG. 4 Alcian blue staining of wild-type and *nar* mutant embryos. *a*, Schematic representation of the ventral view of a wild-type embryo. *b*, *c*, Ventral view of stained wild-type and mutant embryos, respectively, on day 5 of development. The absence of the cartilage of the branchial arches in the mutant is apparent, and malformed pieces of the jaw arches are present (arrow).



normally expressed. Considering also the tight linkage between these two insertions and the associated mutations, we believe that the insertional disruption of these genes is responsible for the observed phenotypes, although definitive proof will require rescue of the mutant phenotypes by addition of functional copies of the disrupted genes.

The gene mutated by insertion 38M is of particular interest because its disruption results in defects in the head, and in particular the pharyngeal arches, a set of structures unique to vertebrates and derived in large part from the cells of the neural crest. We have called this gene *no arches* (*nar*), and have initiated studies to address its role in vertebrate development. Initial observations of *nar* mutant embryos reveal that, although they appear normal during the first 2 days of development, by day 3 they have smaller heads and eyes than their wild-type siblings, and the growth of their pectoral fins is stunted. On days 4 and 5 the difference in head size between mutant and wild-type embryos becomes more extreme, and the pharyngeal arches are largely absent in the mutant. Not all head structures are equally affected, however, as the otic vesicles in the mutant are comparable in size to those in the wild type (Fig. 1*a, b*). Furthermore, no obvious defects are present in the trunk and the tail of mutant embryos, although the development of the liver and gut is slightly retarded.

To examine the structure of the head skeleton, day-5 mutant and wild-type embryos were treated with alcian blue, which stains cartilaginous structures¹². Staining of wild-type embryos shows the seven pharyngeal arches: two jaw arches, the mandibular and the hyoid, and five branchial arches (Fig. 4*a, b*). Also evident is the cartilage of the ethmoid plate, the most anterior portion of the neurocranium. In mutant embryos the cartilage of the branchial arches is largely or entirely absent, and that of the jaw arches is absent or malformed (Fig. 4*c*). The cartilage of the neurocranium is present but reduced in size and the ethmoid plate extends far less anteriorly than in wild-type embryos. Interestingly, the cartilage of both the pharyngeal arches and the ethmoid plate is derived from the neural crest.

A *nar* cDNA sequence of length 1,879 base pairs was obtained from an adult zebrafish cDNA library (provided by D. Grunwald). This fish sequence was compared to the cDNA sequence of the homologous *Drosophila* zinc-finger gene mentioned above and the putative coding sequences were aligned (Fig. 5). The fly gene, *clp*, contains seven zinc-finger motifs, and has been shown biochemically to encode a new type of endoribonuclease that specifically cleaves RNA hairpins⁵. Six of the seven zinc fingers present in the *clp* protein (Clp) are conserved in the putative *nar* protein (Nar) including the five at the amino-terminus which possess the nuclease activity in Clp (Fig. 5). The conserved carboxy-terminal zinc finger is similar in structure to those of retroviral nucleocapsid

proteins which bind to the single-stranded RNA genome of retroviruses¹³. The presence of an RNA-binding motif at the C terminus of Nar, in conjunction with a putative ribonuclease at the N terminus, indicates that the Nar protein may function during embryogenesis by binding to and cleaving RNA.

To examine the temporal pattern of expression of *nar* during embryogenesis, a northern blot analysis was performed (Fig. 6). Two different *nar* transcripts of roughly 1.9 kb and 1.2 kb were detected in unfertilized eggs, indicating that these messages are maternally supplied. During embryogenesis only the 1.9-kb transcript was detected, and was found to be present during gastrulation, neurulation and somitogenesis (6–24 h). The transcript level increases significantly during this period and remains high through 48 h, but then decreases by 96 h. The timing of *nar* expression during embryogenesis is consistent with a role in the formation of the pharyngeal skeleton, which normally takes place during days 2 and 3 of development¹⁴. Expression of *nar* in adults was also examined, and very high levels of both the 1.9-kb, and the 1.2-kb transcripts were detected in the ovaries, whereas low levels of the 1.9-kb transcript were detected in males and in females that had their ovaries removed (data not shown).

Our pilot insertional mutagenesis screen in zebrafish using a retroviral vector has identified three insertional mutants. In two of these, we have found that the proviruses have inserted into and

Zebrafish	MQELIATVDHIKFDLEIAVEQQLGAQPLFPFGMDKSGAAVEYFNRRAA--	48
<i>Drosophila</i>	MDILLANVSGLOPKAERDLIEQVGATPLFPFGMDKSTAAVCFNFTIRNGQE	50
Zebrafish	CMKGGMCFFRRIISGKETVYCKHNLRLGLCKKGDCEFLHEYDMTKMPECYF	98
<i>Drosophila</i>	CDKGSACFFRRIIRGDRITVCKHNLRLGLCKKGDCEFLHEYDMTKMPECYF	100
Zebrafish	YSKFGECNNKCEPFLHIDFESKIKDLPYDRGFCRKGPDCEHRRHTRRVIC	148
<i>Drosophila</i>	YSRFNACHNKECPFLHIDFESKIKDLPYDRGFCRKGPDCEHRRHTRRVIC	150
Zebrafish	VNYLVGFCPEGKSCFKFMHPRFELPMGAT-EOPPEPQOVOT-----QOK	190
<i>Drosophila</i>	MDVLAGFCPEAPSCCKHMHPRFELPLAELGKDLHKKLPTCHYCGELGHR	200
Zebrafish	----QQNMQPI-NRSSQSLIQLTNPNIS--NNNHQRIPNAVGI---VHSN	230
<i>Drosophila</i>	ANSCKQYVGSLEHRRNNINAMDHSGHSGGYSGHSHGIEGADDMQSNHESQ	250
Zebrafish	SHMGG-ARGPRPLDQVTCYKCGEKGHYANKCPKGHLAFLFSQ	271
<i>Drosophila</i>	PHGPGFVKVPTPLEEITCYKCGKNGHYANKCPKGHLAFLSNQHSK	296

FIG. 5 Alignment of the predicted zebrafish Nar protein and the *Drosophila* homologue Clp. The amino-acid identity (shaded) and the zinc-finger motifs (boxes) are shown. These proteins have 69% amino-acid identity (152/221) throughout the conserved zinc-finger motifs (residues 1–176 and 227–271). This alignment was obtained using the program GeneWorks (Oxford Molecular Group).

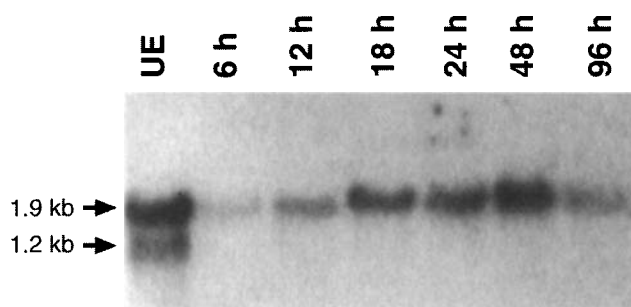


FIG. 6 Northern blot analysis of *nar* expression during embryogenesis. Equivalent amounts of total RNA are present in all lanes as judged by the amounts of 18S rRNA present (not shown). UE, unfertilized egg.

disrupted highly conserved genes that are essential for zebrafish embryogenesis. Using this method on a larger scale, it should now be possible to isolate hundreds of insertional mutants in zebrafish and clone the mutated genes. The use of alternative strategies, such as the screening of haploids or gynogenotes¹⁵, or the development of a retroviral gene trap^{16,17} for use in zebrafish, could make the isolation of thousands of insertional mutants possible. □

Methods

Southern and northern blots. Genomic DNA was isolated after lysis at 55 °C in 100 mM Tris buffer (pH 8.0), 5 mM EDTA, 200 mM NaCl, 0.4% SDS and 100 µg ml⁻¹ proteinase K. Total RNA was prepared using the RNazol B method (Tel-Test, Friendswood, Texas). All gels were transferred to Hybond N+ membranes (Amersham), and filters were hybridized and washed as described⁶.

Reverse transcriptase PCR. First-strand cDNA synthesis was done at 42 °C for 1 h using 1 µg of total RNA. PCR was done in 67 mM Tris buffer (pH 8.8), 16.6 mM (NH₄)₂SO₄, 10 mM β-mercaptoethanol, 170 µg ml⁻¹ bovine serum albumin, 2 mM MgCl₂, with 0.2 µM of each dNTP, and 0.4 µM of each primer. The PCR sequence used was as follows: 30 cycles of 15 s at 94 °C, 30 s at 55 °C, and 45 s at 72 °C, with an initial 2 min denaturation step and a final 5 min elongation step.

Alcian-blue staining. Embryos were fixed in 4% paraformaldehyde for 2 h at room temperature, and then stained overnight with 0.1% alcian blue in 90% ethanol/10% glacial acetic acid. Embryos were rehydrated into phosphate buffered saline (PBS), treated for 1–2 h in a 1% trypsin solution saturated with sodium borate, and were bleached in 1% KOH (w/v), 3% H₂O₂, and 100 mM NaCl. Stained embryos were then washed with PBS and cleared with glycerol.

Received 10 June; accepted 16 September 1996.

- Haffter, P. et al. *Development* **123**, 1–36 (1996).
- Driever, W. et al. *Development* **123**, 37–46 (1996).
- Cooley, L., Kelley, R. & Spradling, A. *Science* **239**, 1121–1128 (1988).
- Bier, E. et al. *Genes Dev.* **3**, 1273–1287 (1989).
- Bai, C. & Tolia, P. *Mol. Cell. Biol.* (in the press).
- Gaiano, N., Allende, M., Amsterdam, A., Kawakami, K. & Hopkins, N. *Proc. Natl Acad. Sci. USA* **93**, 7777–7782 (1996).
- Burns, J. C. et al. *Proc. Natl Acad. Sci. USA* **90**, 8033–8037 (1993).
- Hopkins, N. *Proc. Natl Acad. Sci. USA* **90**, 8759–8760 (1993).
- Lin, S. et al. *Science* **265**, 666–669 (1994).
- Meisler, M. H. *Trends Genet.* **8**, 341–344 (1992).
- Altschul, S., Gish, W., Miller, W., Myers, E. & Lipman, D. J. *Mol. Biol.* **215**, 403–410 (1990).
- Dingerkus, G. & Uhler, D. L. *Stain Tech.* **32**, 229–231 (1977).
- Summers, M. F. J. *Cell. Biochem.* **45**, 41–48 (1991).
- Schilling, T. F. & Kimmel, C. B. *Development* **120**, 483–494 (1994).
- Westerfield, M. (ed.) *The Zebrafish Book* (Univ. of Oregon Press, Eugene, 1993).
- Friedrich, G. & Soriano, P. *Genes Dev.* **5**, 1513–1523 (1991).
- Chen, J. et al. *Meth. Mol. Genet.* **4**, 123–140 (1994).

ACKNOWLEDGEMENTS. We thank D. Thompson, E. Abrams and B. Hayward for maintaining the fish, M. Elrod-Erickson for helpful discussion and P. Tolia for sharing data before publication. We also thank C. Nüsslein-Volhard for helpful discussions and R. Bosselman for suggestions and support. This work was funded by grants to N.H. from Amgen, the NIH, the Human Frontier Science Program, and a core grant from the NCI to the Center for Cancer Research at MIT. Additional funding was provided by fellowships from the HHMI (A.A.), NIH (M.A.), the Yamada Science Foundation and the Toyobo Biotechnology Foundation (K.K.), and the Damon Runyon-Walter Winchell Foundation (T.B.).

CORRESPONDENCE and requests for material should be addressed to N.H. (e-mail: nhopkins@mit.edu). The cDNA sequence of the zebrafish *nar* gene has been deposited in Genbank, accession number U70479.

Partnership between DPC4 and SMAD proteins in TGF-β signalling pathways

Giorgio Lagna*, Akiko Hata†, Ali Hemmati-Brivanlou* & Joan Massagué†

† Memorial Sloan-Kettering Cancer Center and Howard Hughes Medical Institute, * The Rockefeller University, Laboratory of Molecular Embryology, New York, New York 10021, USA

THE TGF-β/activin/BMP superfamily of growth factors signals through heteromeric receptor complexes of type I and type II serine/threonine kinase receptors¹. The signal originated by TGF-β-like molecules appears to be transduced by a set of evolutionarily conserved proteins known as SMADs, which upon activation directly translocate to the nucleus where they may activate transcription². Five SMAD proteins have so far been characterized in vertebrates¹. These factors are related to the mediator of *decapentaplegic* (*dpp*) signalling, *mothers against dpp* (*Mad*), in *Drosophila*³ and to the *Sma* genes from *Caenorhabditis elegans*⁴. *Smad1* and *Smad2* have been shown to mimic the effects of BMP and activin, respectively, both in *Xenopus* and in mammalian cells^{2,5–9}, whereas *Smad3* (a close homologue of *Smad2*) and the related protein DPC4, a tumour-suppressor gene product¹⁰, mediate TGF-β actions¹¹. We report here that DPC4 is essential for the function of *Smad1* and *Smad2* in pathways that signal mesoderm induction and patterning in *Xenopus* embryos, as well as antimitogenic and transcriptional responses in breast epithelial cells. DPC4 associates with *Smad1* in response to BMP and with *Smad2* in response to activin or TGF-β. DPC4 is therefore a regulated partner of SMADs that function in different signalling pathways of the TGF-β family.

In mammalian cells, transforming growth factor (TGF)-β and activin, signalling through distinct but highly homologous receptors^{12,13}, can trigger similar effects, including cell-cycle arrest and activation of the plasminogen-activator inhibitor-1 (PAI-1) gene promoter¹². In *Xenopus*, mesoderm induction involves endogenous activin-like signals¹⁴ and can be triggered by incubation of ectodermal explants with activin or the related Vg1 protein^{15–18}. Bone morphogenetic proteins (BMPs) play a central role in embryonic patterning. In *Xenopus*, BMP4 and related factors have been implicated in the expression of ventral mesodermal genes and in the repression of neural fate¹⁹. To study the potential involvement of DPC4 in these pathways, we first verified that a DPC4 homologue is expressed in early embryos. We isolated a complementary DNA encoding the carboxy-terminal domain of a typical SMAD protein that is 89% identical to the corresponding sequence in human DPC4 (ref. 10) and only 40–46% similar to human or *Xenopus* *Smad1* and *Smad2* (ref. 5, 7) (Fig. 1A, and data not shown). We conclude that this cDNA corresponds to a *Xenopus* homologue of DPC4 and refer to it as *XSmad4*. As previously shown for *XSmad1* and *XSmad2* (refs 5, 7), quantitative polymerase chain reaction coupled with reverse transcription (RT-PCR) and whole-mount *in situ* hybridization of staged embryos showed that *XSmad4* transcripts are also present in the maternal RNA pool and are ubiquitously expressed at least until the neural-groove stage (data not shown). To investigate a possible role for DPC4 *in vivo* in gain-of-function experiments, we used a full-length human *Smad4*/DPC clone². *In vitro* synthesized DPC4 transcripts were injected into the animal pole of two-cell *Xenopus* embryos. Injected embryos were allowed to develop until the blastula stage, when ectodermal explants (animal caps) were excised and cultured until sibling controls reached midgastrula stages (Fig. 1B). In this assay, animal caps derived from uninjected embryos differentiate as epidermis. Animal caps expressing low levels of *Smad2*, however, give rise to ventral

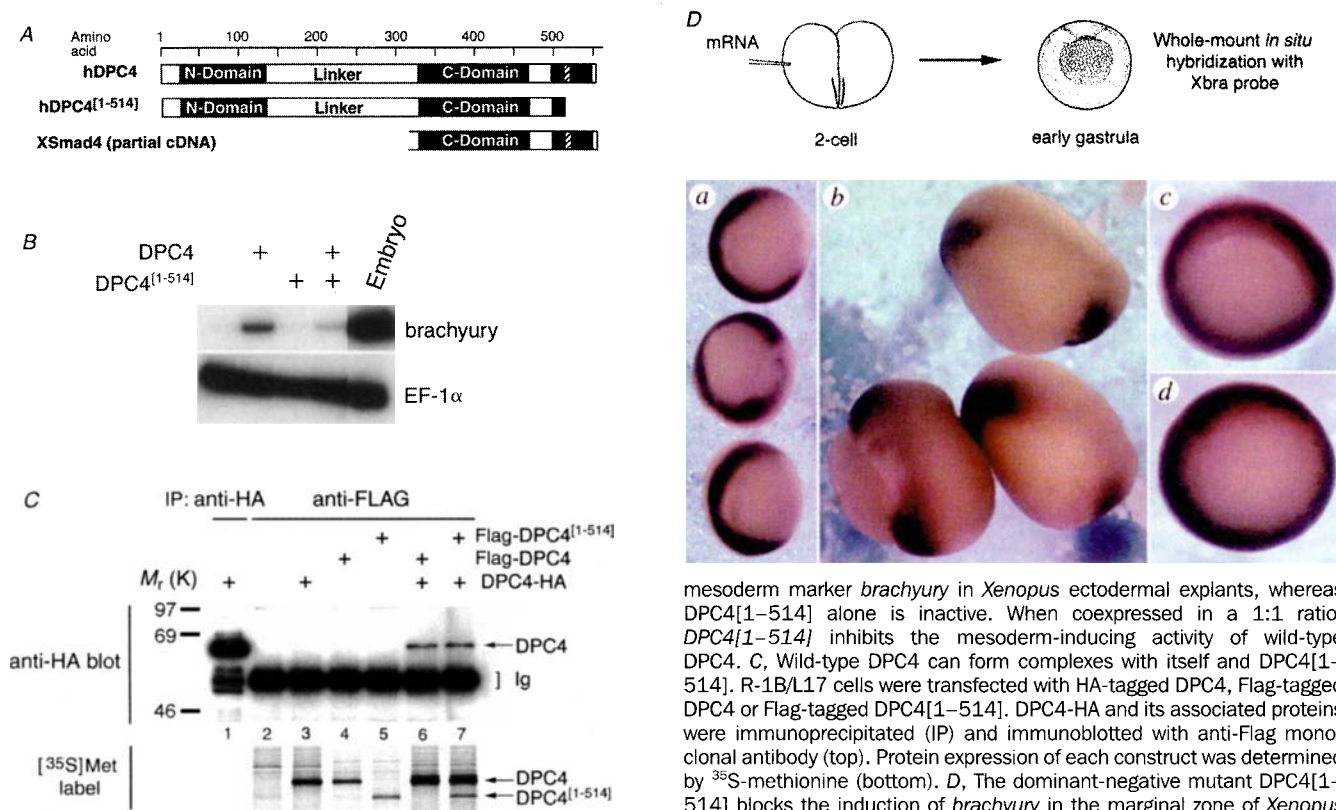


FIG. 1 Human DPC4 acts as a mesoderm inducer and the dominant-negative mutant DPC4[1–514] blocks mesoderm induction *in vivo*. *A*, The domain structure of wild-type human DPC4, the naturally occurring mutant DPC4[1–514], and the C-terminal domain of the *Xenopus* homologue Smad4 (XSmad4). The conserved N- and C-terminal domains (black) and the mutational hot spot (hatched) are indicated. *B*, DPC4 induces the

mesoderm, whereas at higher doses, *Smad2* also induces dorsal mesodermal markers^{5,8}, mimicking the dose-dependent effects of activin addition. On the other hand, *Smad1* overexpression mimics BMP signalling in that it ventralizes dorsal mesoderm^{2,5,7,8}. Figure 1B shows that injection of *DPC4* transcripts leads to induction of the general mesoderm marker *brachyury* (*Xbra*)²⁰. As little as 250 pg *DPC4* RNA was sufficient to induce this gene (data not shown). *DPC4* also induced the expression of the early ventral marker *Xwnu8* (Fig. 2a), but, unlike *Smad2*, it failed to induce the dorsal markers *gooseoid* (Fig. 2b), *chordin* or *noggin* (data not shown). These results indicate that *DPC4* alone can act as a ventral mesoderm inducer in the context of animal cap explants, thus mimicking the effect of low concentrations of

activin^{15,16}

Many *DPC4* mutations in human pancreatic carcinoma and other malignancies occur in a short region within the conserved C-terminal domain¹⁰ (Fig. 1A). One such nonsense mutation found in a pancreatic carcinoma generates DPC4[1-514], a protein with a small C-terminal truncation¹⁰ (Fig. 1A). This truncation disrupts the activity of DPC4 in a transcriptional activation assay². Consistently, *DPC4*[1-514] failed to induce mesoderm in the animal cap assay (Fig. 1B). Furthermore, mesoderm induction by *DPC4* was inhibited by co-injection of an equal amount of *DPC4*[1-514] (Fig. 1B). To determine the basis of this potent inhibitory effect, we investigated whether DPC4 could form homo-oligomers and whether DPC4[1-514] might be incorporated into such com-

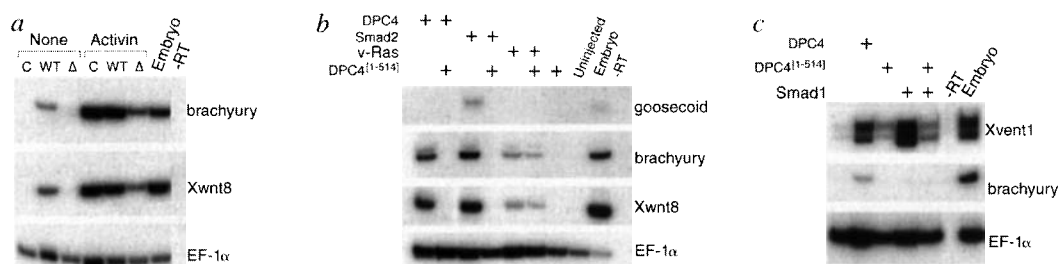


FIG. 2 Requirement of DPC4 for activin, Smad2 and Smad1 function. *a*, The overexpression of DPC4[1–514] (Δ) in animal caps impairs induction of the mesodermal markers *brachyury* and *Xwnt8* by activin. Wild-type DPC4 (WT), which by itself induces mesoderm, does not interfere with activin function. C, uninjected caps. *b*, DPC4[1–514] blocks expression of

dorsal and ventral mesodermal markers induced by Smad2, but does not affect the ability of v-Ras to induce mesoderm. c, Induction of the ventral marker *Xvent1* by Smad1 is impaired by coexpression of DPC4[1–514]. Consistent with previous reports, Smad1 fails to induce *brachyury*.

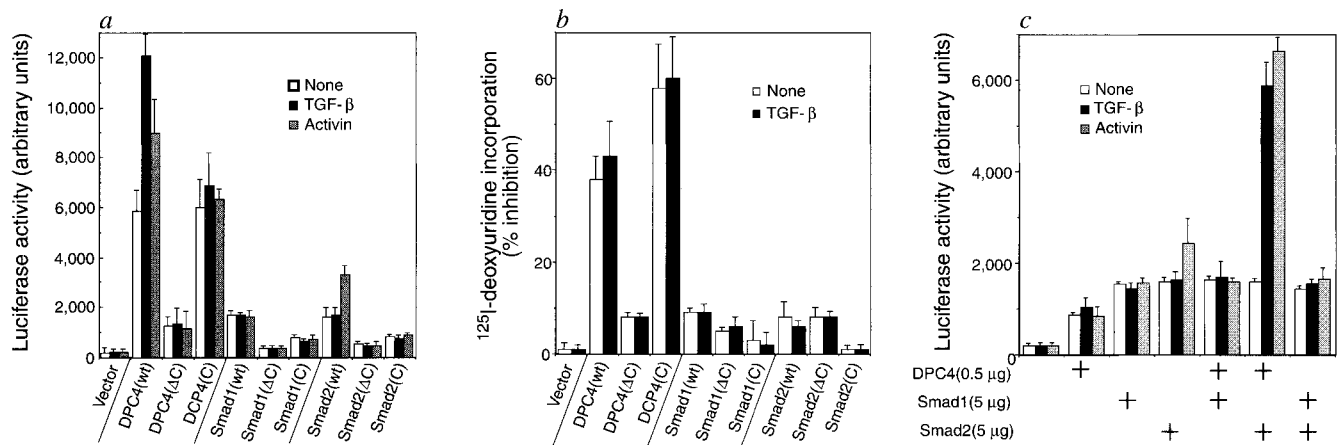


FIG. 3 DPC4, together with Smad2, mediates activin/TGF- β signals in a DPC4-null cell line (MDA-MB468). **a**, DPC4, but not Smad1, mediates transcriptional activation of the *PAI-1* promoter by TGF- β and activin. The human breast-carcinoma epithelial cell line MDA-MB468 was transiently transfected with 5 μ g full-length SMADs (WT), C-terminally truncated SMAD mutants (Δ C) or SMAD C-terminal domain (C) constructs with the 3TP-lux vector. Unlike wild-type DPC4, the C-terminal truncation mutant DPC4[1–514] was unable to induce expression of the reporter gene. The DPC4 C-terminal domain alone was as active as wild-type DPC4 in the induction of the luciferase reporter, but it failed to confer responsiveness to TGF- β or

activin. Smad2 alone weakly mediated activin induction, but Smad1 had no effect. **b**, Full-length DPC4 and its C-terminal domain can by themselves inhibit growth of MDA-MB468 cells in a TGF- β -independent manner. The difference in the amount of inhibition by DPC4(WT) and DPC4(C) reflects differences in transfection efficiency (as measured by an internal control). Smad1 and Smad2 alone did not arrest cell growth. **c**, Synergistic effect of Smad2 and DPC4 in the activation of the *PAI-1* promoter. A small amount of DPC4 DNA was insufficient by itself to confer responsiveness to TGF- β and activin in the 3TP-lux assay; in conjunction with Smad2 DNA, however, DPC4 rescued the ability of MDA-MB468 cells to respond to both activin and TGF- β .

plexes. We therefore co-transfected the mink-lung epithelial cell line R-1B/L17 (ref. 21) with two different DPC4 constructs, one tagged with the Flag epitope and the other tagged with haemagglutinin (HA) epitope. Immunoprecipitation of cell lysates with anti-Flag antibody and immunoblotting of the precipitates with anti-HA antibody revealed an association between the two tagged DPC4 products (Fig. 1C, lane 6). Furthermore, HA-tagged DPC4 was still able to form a complex with co-transfected Flag-tagged (Fig. 1C, lane 7). These results indicate that DPC4 specifically associates with itself or with DPC4[1–514] in a ligand-independent manner, providing a possible explanation for the dominant-negative effect of this mutant allele in *Xenopus*.

The availability of a dominant-negative DPC4 construct allowed us to investigate the potential requirement of endogenous DPC4 function in mesoderm induction. Mesoderm induction is manifested by the appearance of *brachyury* transcript in the equatorial zone of the early gastrula²⁰ (Fig. 1D, c). Injection of DPC4[1–514] RNA into the equatorial region of one blastomere of two-cell embryos prevented *brachyury* expression on one side of the embryo as development reached the early gastrula stage (Fig. 1D, a and b), an effect that was rescued by co-injection of wild-type DPC4 (Fig. 1D, d). These results indicate the potential requirement for endogenous DPC4 in mesoderm induction.

To test for DPC4 involvement in activin signalling, DPC4[1–514] was injected into the animal pole of fertilized *Xenopus* eggs. Animal caps derived from these embryos showed reduced formation of mesoderm in response to activin (Fig. 2a, lane Δ). Moreover, as Smad2 is thought to mediate dorsal mesoderm formation by activin^{5,8}, we tested for a functional interaction between Smad2 and DPC4 in this process. Consistent with such an interaction, mesoderm induction by Smad2 was completely inhibited by co-injection of DPC4[1–514] (Fig. 2b, lanes 3 and 4). DPC4[1–514] also inhibited the induction of a secondary axis by Smad2 (ref. 8, and data not shown). Mesoderm induction by *v-ras*, which is thought to occur through the fibroblast growth factor (FGF)/Ras/MAP kinase pathway²², was found not to be susceptible to the dominant-negative effect of DPC4[1–514] (Fig. 2b, lanes 5 and 6). These observations support a model in which a requirement for endogenous DPC4 function in mesoderm induction is specific for the activin/Smad2 pathway.

We next investigated whether the action of the BMP mediator Smad1 was also affected by DPC4[1–514]. Expression of *Xenopus* Smad1 in animal caps induces expression of an array of ventral mesodermal markers such as *Xwnt8* and *globin*^{5,7}, but not *brachyury*⁵. Here Smad1 is shown to induce the expression of another ventral marker, *Xvent1* (ref. 23), without inducing *brachyury* (Fig. 2c). DPC4 injection also induced the expression of *Xvent1* (Fig. 2c). Surprisingly, co-injection of DPC4[1–514] inhibited *Xvent1* induction by Smad1 (Fig. 2c). Thus, DPC4[1–514] antagonizes Smad1 in *Xvent1* induction in just the way it interferes with mesoderm induction by Smad2.

As DPC4 seemed to be involved with Smad1 in BMP signalling and with Smad2 in activin signalling we investigated the involvement of DPC4 in TGF- β signalling in a DPC4-defective cell line. In contrast to non-tumorigenic mammary epithelial cell lines which respond strongly to TGF- β , the human breast-carcinoma cell line MDA-MB468, which has a DPC4 homozygous deletion²⁴, did not respond to TGF- β or to activin in either of two assays, activation of the *PAI-1* gene promoter (3TP-lux reporter)¹² (Fig. 3a) or growth inhibition (Fig. 3b). Transient transfection of wild-type DPC4 into MDA-MB468 cells strongly increased the basal level of 3TP-lux expression, whereas transfection of DPC4[1–514] did not ([DPC4(Δ C)] in Fig. 3a). Furthermore, DPC4 conferred responsiveness to both activin and TGF- β , because both factors increased 3TP-lux expression further (Fig. 3a). Transfection of a construct encoding the C-terminal domain of DPC4 [DPC4(C)], which is constitutively active in a transcription assay², activated 3TP-lux expression, but this effect was not enhanced by activin or TGF- β addition (Fig. 3a). Smad1 and Smad2 tested as either full-length or C-terminal domains had only very small effects on 3TP-lux expression, although this was significantly increased by activin in cells transfected with wild-type Smad2 ([Smad2(wt)] Fig. 3a). These results indicate that, when overexpressed in MDA-MB468 cells, DPC4 or its biologically active domain can mimic the action of activin and TGF- β and, in the case of wild-type DPC4, can confer responsiveness to these factors. DPC4 overexpression also inhibited growth of these cells (Fig. 3b) to an extent that was related to the proportion of cells expressing transfected DNA, as determined by co-transfection of green fluorescence protein as an internal control (data not shown). This growth inhibition by DPC4

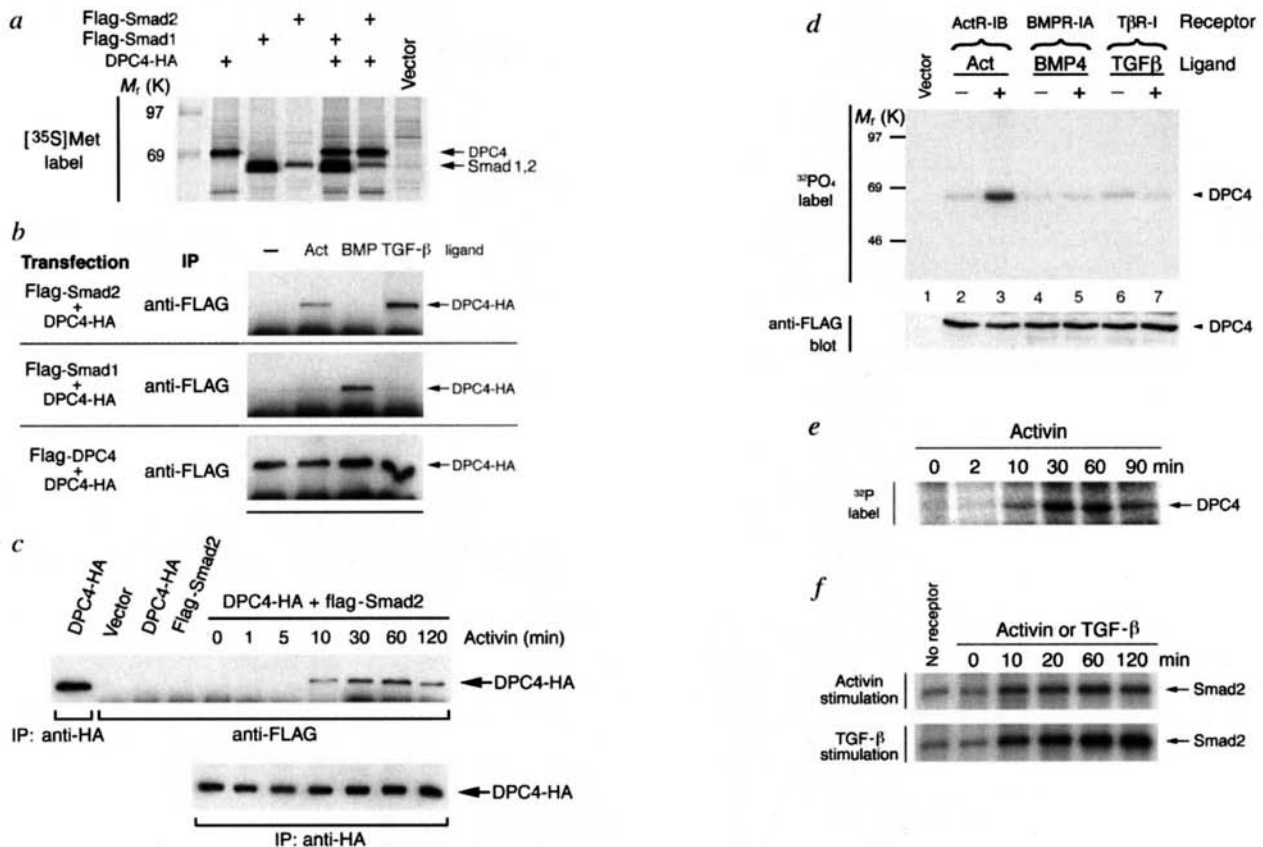


FIG. 4. Ligand-dependent phosphorylation and interaction of DPC4 with Smad1 and Smad2. *a*, Expression of tagged DPC4 and SMAD proteins was monitored in ³⁵S-methionine-labelled cells. *b*, Ligand-induced interaction between DPC4 and SMADs. Cells transiently transfected with the indicated constructs were stimulated for 1 h with activin A (Act), BMP4 (BMP) or TGF- β . Flag-tagged SMADs and their associated proteins were immunoprecipitated by anti-Flag antibody and immunoblotted using anti-HA antibody. DPC4-Smad2 complexes formed only in response to TGF- β and activin, whereas DPC4-Smad1 complexes were only detected upon stimulation with BMP. *c*, Kinetics of interaction between Smad2 and DPC4 after

stimulation by activin. Formation of the DPC4-Smad2 complex was detected as early as 10 min after stimulation and was stable for at least 2 h. *d*, Phosphorylation of DPC4 by activin. R-1B/L17 cells were transiently transfected with either vector alone (lane 1) or with Flag-DPC4 DNA and the indicated type-I receptors (lanes 2–7). ActR-IB (type II receptor) was transfected with BMPR-IA in lanes 4 and 5. Phosphorylated (32 P-labelled) and total (anti-Flag blot) DPC4 are shown. *e*, Kinetics of DPC4 phosphorylation by activin. *f*, Smad2 can be phosphorylated by both activin and TGF- β . Flag-Smad2 was transfected into R-1B/L17 cells with ActR-IB or β IR-I and stimulated with activin A or TGF- β , respectively, for the indicated times.

or DPC4[1-514] was not increased by addition of TGF- β , possibly because of DPC4 overexpression (Fig. 3b).

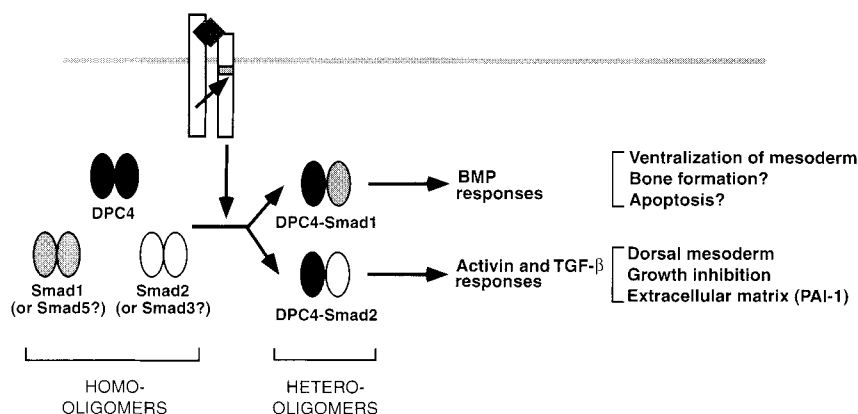
The ability of DPC4 to mediate TGF- β and activin responses in combination with Smad2 was established in MDA-MB468 cells by transfection of small amounts of *DPC4* DNA (Fig. 3c). Under these conditions, transfection of *DPC4* alone had only a small effect on 3TP-lux induction and did not confer responsiveness to TGF- β or activin, although co-transfection of *DPC4* and *Smad2* conferred a strong responsiveness to both factors (Fig. 3c). Co-transfection of *Smad1* and *DPC4*, or of *Smad1* and *Smad2*, did not have this effect. Taken together, these results indicate that DPC4 and Smad2 act together in activin and TGF- β signalling pathways in breast epithelial cells.

We next investigated whether these interactions could involve an association between DPC4 and SMAD proteins. Like DPC4 (Fig. 1C), Smad1 and Smad2 can form homomeric complexes (data not shown). We transfected R-1B/L17 cells with HA-tagged DPC4, Flag-tagged Smad1 or Flag-tagged Smad2, either singly or in binary DPC4, Smad combinations. Although these proteins were all expressed in the transfected cells (Fig. 4a), there was no association between Smad1 or Smad2 and DPC4 in the absence of exogenous ligands (Fig. 4b). However, Smad2 and DPC4 did form a complex when the cells were stimulated with activin or TGF- β : this complex was detectable within 10 min of addition of activin and was stable for at least 2 h (Fig. 4c). The DPC4–Smad2

interaction was ligand specific as it did not occur in cells stimulated with BMP4 (Fig. 4b). Conversely, Smad1 and DPC4 formed a complex when the cells were stimulated with BMP4 but not with activin or TGF- β (Fig. 4b). The interaction of DPC4 with itself in the absence of exogenous ligands was not affected by activin, TGF- β or BMP (Fig. 4b). Therefore, SMADs mediating the action of particular TGF- β family members are specifically induced to associate with DPC4 in response to these factors.

Smad1 becomes phosphorylated in response to BMP^{2,6,25}, whereas Smad2 is phosphorylated in response to TGF- β (ref. 9). As phosphorylation appears to be a feature of SMAD activation, we tested the receptors for activin, BMP or TGF- β for their ability to mediate DPC4 phosphorylation. R-1B/L17 cells were used because they respond to activin, TGF- β and BMP after transfection of the appropriate receptors^{12,26}. Activin addition to cells transfected with the activin type-I receptor (ActR-IB) induced phosphorylation of co-transfected DPC4 (Fig. 4d) within 10 min and lasted for at least 90 min (Fig. 4e). DPC4 phosphorylation was not detectable in cells stimulated through the TGF- β or BMP receptors (Fig. 4d); addition of activin to ActR-IB-transfected cells or of TGF- β to T β R-I-transfected cells induced a limited but significant increase in Smad2 phosphorylation⁹ (Fig. 4f). This concomitant phosphorylation, oligomerization and nuclear translocation of DPC4/SMAD suggests that they may be causally linked events.

FIG. 5 Proposed SMAD function in TGF- β family signalling pathways. In response to ligand activation of two consecutive receptor kinases, a shared SMAD, DPC4 (ref. 10), associates with the pathway-specific Smad1 (refs 2, 5–7) or Smad2 (refs 5, 8, 9) or their isoforms Smad5 (ref. 25) or Smad3 (ref. 11), respectively. Each SMAD–DPC4 combination leads to a specific set of responses.



The features of SMAD proteins have suggested a model in which each member of the TGF- β family signals through its own specific SMAD: thus BMP acts through Smad1, and activin through Smad2 (refs 2, 5–9, 25). Smad3, an isoform of Smad2, participates in TGF- β signalling with DPC4 (ref. 11). Smad1 and its close homologue Smad5, as well as Smad2 and Smad3, form a cluster of closely related gene products^{11,25}; DPC4 is structurally more distant. By analysing the pathways of activin, BMP and TGF- β signalling in different systems (*Xenopus* embryos and mammalian cells), we have shown that all the responses tested depend on interaction between DPC4 and one of the other SMADs. Functional association between SMADs had already been inferred from genetic evidence in *C. elegans*, in which the *Sma-2*, *Sma-3* and *Sma-4* genes are all individually required for *daf-4* signalling¹. Our results indicate that DPC4 plays a central role in these interactions, acting in partnership with different SMADs in distinct TGF- β signalling pathways (Fig. 5). The entire SMAD signalling network may therefore be disabled in cancer cells that have lost DPC4 function. □

Methods

Isolation of XSmad4 cDNA. The following degenerate polymerase chain reaction (PCR) primers complementary to conserved regions in human DPC4 (ref. 10) and *C. elegans* Sma-4 (ref. 4) were used: 5'-TGGTGTCTCATHGCITATY-WYGA-3' and 5'-TCRATCCARCAIGGTYTCYTDTAT-3'. A fragment amplified from tadpole cDNA was used to screen a *Xenopus* oocyte cDNA library (gift from P. Klein) and isolate the partial XSmad4 clone.

Animal cap assay and whole-mount *in situ* hybridization. DPC4 RNA was synthesized *in vitro* in the presence of cap analogue using the mMessage mMachine kit (Ambion). The DPC4 DNA template was generated by subcloning the human cDNA into the EcoRI/BglII sites of pSP64TEN (P. Wilson, unpublished). The human Smad2 cDNA (F. Liu and J.M., unpublished) was inserted in the EcoRI/XhoI sites of the pCS2 vector (gift from L. Turner). *v-ras* RNA was transcribed from the EcoRI-linearized 64T-ras-VR plasmid (constructed by M. Whitman). Human Smad1 cDNA² was subcloned into the EcoRI/BglII sites of pSP64TEN. All the constructs were transcribed *in vitro* with SP6 RNA polymerase. Where indicated, injected and untreated animal caps were incubated with 100 pM activin in 0.5 $\times$ MMR medium supplemented with 0.5% BSA. For the animal cap assay, RNA (10 nl, 0.25–4 ng) was introduced in the animal pole of two-cell *Xenopus* embryos. Animal caps were explanted at blastula stage (stage 8–9) and cultured to gastrula stage (stage 11–12). Total RNA from the isolated explants and control sibling embryos was extracted with RNazol B (TEL-TEST, Inc.). RT-PCR has been described²⁷, as have primers for *brachyury*, *Xwnt8*, *goosecoid* and *EF-1 α* (ref. 27). The *Xwnt1* (ref. 23) primers were: 5' to 3' (AGG CAA CAG ATG GAA AGG AC); (CGC TGG AGA TCA GAT TTG G). Whole-mount *in situ* hybridization was done on albino embryos as described²⁸. The antisense probe (gift from P. Wilson) was synthesized in the presence of digoxigenin-11-UTP (Boehringer Mannheim).

Transfection, metabolic labelling, immunoprecipitation and western blot. R-1B/L17 and MDA-MB468 cells (ATCC HTB 132) were transfected with the indicated constructs and metabolically ³⁵S- or ³²P-labelled as described²⁹. Two days after transfection, cells were incubated in the presence or absence of 5 nM activin A (Genentech), 10 nM BMP4 (gift from V. Rosen), or 100 pM TGF- β for 1 h or as indicated. Transfected cells were lysed in TNE buffer (10 mM Tris, pH 7.8, 150 mM NaCl, 1 mM EDTA, 1% Nonidet P-40) containing protease

inhibitors and immunoprecipitated with anti-Flag M2 monoclonal antibody (IBI; Eastman Kodak) or anti-HA monoclonal antibody 12CA5 (Boehringer Mannheim). Immunoprecipitates were separated by SDS-PAGE and transferred to Immobilon-P membranes (Millipore). Tagged proteins were detected by western blotting, followed by chemiluminescence (ECL, Amersham).

Luciferase and growth-inhibition assays. Cells were transiently transfected with 5 μ g (unless otherwise indicated) of SMAD constructs and 3 μ g 3TP-lux plasmid, incubated in the presence of ligands for 18 h and collected. Luciferase activity was assayed in a Promega system. Results from 3TP-lux assays are the mean $\pm$ s.d. of three experiments. For the growth-inhibition experiment, MDA-MB468 cells were transiently transfected with the indicated constructs, incubated in the presence or absence of 100 pM TGF- β 1 for 20 h, and labelled with 2 μ Ci ml⁻¹ ¹²⁵I-iododeoxyuridine for the last 3 h. ¹²⁵I-iododeoxyuridine incorporation of non-transfected cells cultured in the absence of TGF- β was set at 100%. Data are plotted as the percentage of inhibition of ¹²⁵I-iododeoxyuridine incorporation in transfected cells. A vector driving the expression of green fluorescence protein (0.5 μ g, pEGFP-C3, Clontech) was co-transfected to score transfection efficiency. Data shown are the mean $\pm$ s.d. of at least three experiments.

Received 9 September; accepted 3 October 1996.

- Massagué, J. *Cell* **85**, 947–950 (1996).
- Liu, F. et al. *Nature* **381**, 622–623 (1996).
- Sekelsky, J. J., Newfield, S. J., Raftery, L. A., Chartoff, E. H. & Gelbart, W. M. *Genetics* **139**, 1347–1358 (1995).
- Savage, C. et al. *Proc. Natl Acad. Sci. USA* **93**, 790–794 (1996).
- Graff, J. M., Bansal, A. & Melton, D. A. *Cell* **85**, 479–487 (1996).
- Hoodless, P. A. et al. *Cell* **85**, 489–500 (1996).
- Thomsen, G. *Development* **122**, 2359–2366 (1996).
- Baker, J. & Harland, R. M. *Genes Dev.* **10**, 1880–1889 (1996).
- Eppert, K. et al. *Cell* **86**, 543–552 (1996).
- Hahn, S. A. et al. *Science* **271**, 350–353 (1996).
- Zhang, Y., Feng, X.-H., Wu, R.-Y. & Derynck, R. *Nature* **383**, 168–172 (1996).
- Cárcamo, J. et al. *Mol. Cell. Biol.* **14**, 3810–3821 (1994).
- ten Dijke, P. et al. *Oncogene* **8**, 2879–2887 (1993).
- Hemmati-Brivanlou, A. & Melton, D. A. *Nature* **359**, 609–614 (1992).
- Smith, J. C., Price, B. M. J., Van Nimmen, K. & Huylebroeck, D. *Nature* **345**, 729–731 (1990).
- Asashima, M. et al. *Roux's Arch. Dev. Biol.* **198**, 330–335 (1990).
- Sokol, S., Wong, G. G. & Melton, D. A. *Science* **249**, 561–564 (1990).
- Thomsen, G. H. & Melton, D. A. *Cell* **74**, 433–441 (1993).
- Hogan, B. L. M. *Genes Dev.* **10**, 1580–1594 (1996).
- Smith, J. C., Price, B. M. J., Green, J. B. A., Weigel, D. & Herrmann, B. G. *Cell* **67**, 79–87 (1991).
- Attisano, L., Wrana, J. L., Montalvo, E. & Massagué, J. *Mol. Cell. Biol.* **16**, 1066–1073 (1996).
- Whitman, M. & Melton, D. A. *Nature* **357**, 252–254 (1992).
- Gawantka, V., Delius, H., Hirschfeld, K., Blumenstock, C. & Niehrs, C. *EMBO J.* **14**, 6268–6279 (1995).
- Schutte, M. et al. *Cancer Res.* **56**, 2527–2530 (1996).
- Yingling, J. M. et al. *Proc. Natl Acad. Sci. USA* **93**, 8940–8944 (1996).
- Liu, F., Ventura, F., Doody, J. & Massagué, J. *Mol. Cell. Biol.* **15**, 3479–3486 (1995).
- Wilson, P. A. & Melton, D. A. *Curr. Biol.* **4**, 676–686 (1994).
- Harland, R. M. in *Methods in Cell Biology* (eds Kay, B. K. & Peng, H. B.) (Academic, San Diego, 1991).
- Wrana, J. L., Attisano, L., Wieser, R., Ventura, F. & Massagué, J. *Nature* **370**, 341–347 (1994).

ACKNOWLEDGEMENTS. G.L. and A.H. contributed equally to this work. We thank F. Liu for providing the human Smad2 cDNA clone before publication, J. Doody for technical assistance, the Genetics Institute for recombinant human BMP4, and P. Klein for the oocyte cDNA library. G.L. thanks F. Carnevali and I. Bozzoni for their encouragement. This work was supported by funds from the Rockefeller University, the Goldsmith foundation and the Searle Scholars Program/The Chicago Community Trust to A.H.B. and grants from the NIH to A.H.B., J.M. and to Memorial Sloan-Kettering Cancer Center. G.L. is a graduate fellow on leave of absence from the University 'La Sapienza' of Rome. A.H. is a research associated and J.M. is an investigator of the Howard Hughes Medical Institute.

CORRESPONDENCE and requests for materials should be addressed to J.M. (e-mail: j-massagué@ski.mskcc.org) or A.H.B. (e-mail: brnlu@rockefeller.edu).

Rapid shuttling of NF-AT in discrimination of Ca^{2+} signals and immunosuppression

Luika A. Timmerman^{*†}, Neil A. Clipstone^{*‡},
Steffan N. Ho^{*}, Jeffrey P. Northrop^{*‡}
& Gerald R. Crabtree^{*}

^{*} The Departments of Developmental Biology and Pathology, and Howard Hughes Medical Institute, Stanford University School of Medicine, and [†] The Program in Immunology, Stanford University, Stanford, California 94305, USA

[‡] Present addresses: Department of Microbiology and Immunology, Northwestern University, Chicago, Illinois 60611, USA (N.A.C.); Affymax Research Institute, 3410 Central Expressway, Santa Clara, California 95051, USA (J.P.N.).

CELLS need to distinguish between transient Ca^{2+} signals that induce events such as muscle contraction, secretion, adhesion and synaptic transmission, and sustained Ca^{2+} signals that are involved in cell proliferation and differentiation. The latter class of events is blocked in lymphocytes by the immunosuppressive drugs cyclosporin A and FK506, which inhibit calcineurin, a Ca^{2+} -activated serine/threonine phosphatase necessary for the nuclear import of NF-AT transcription factors¹⁻⁴. Here we report that sustained high concentrations of Ca^{2+} , but not transient pulses, are required to maintain NF-AT transcription factors in the nucleus, where they participate in Ca^{2+} -dependent induction of genes required for lymphocyte activation and proliferation. Furthermore, overexpression and constitutive nuclear localization of NF-AT, but not Jun, Fos, NF- κ B, Oct or Ets family members, renders the interleukin-2 enhancer in Jurkat T lymphocytes resistant to FK506 and cyclosporin A. Thus a primary effect of these immunosuppressive reagents is to control the subcellular localization of the NF-AT family of transcription factors.

Shortly after the discovery of T-cell activation⁵, a requirement for extracellular calcium^{6,7} during stimulation was defined and correlated with a rapid spike and subsequent sustained rise in intracellular calcium levels ($[\text{Ca}^{2+}]_i$). Studies using pharmacological inhibitors or cells with mutations preventing signal transduction have demonstrated that $[\text{Ca}^{2+}]_i$ must exceed 400 nM for about 2 h for T cells to become committed to activation⁸. One role of calcium is to activate the CsA-sensitive phosphatase calcineurin, which either directly or indirectly leads to the nuclear transloca-

tion of members of the NF-AT transcription factor family^{3,4,9}. Biochemical and gene-ablation studies indicate that NF-AT family members control the induction of early immune-response genes, such as those for interleukin (IL)-2, IL-3, IL-4, granulocyte macrophage colony-stimulating factor, CD40 ligand, and the Fas ligand, which control T-cell proliferation and differentiation, and collectively coordinate the actions of other cells necessary for the immune response¹⁰⁻¹².

To examine potential differences between the effects on NF-AT of transient and sustained changes in $[\text{Ca}^{2+}]_i$, we examined the subcellular localization of NF-AT family members in a variety of established T-cell lines and murine T lymphocytes, using specific antibodies against the three NF-AT family members found in peripheral lymphocytes: NF-ATc, NF-ATp and NF-ATx/c3/4 (refs 13-16). (Note on NF-ATc nomenclature: to maintain the distinction between the cytosolic (C) and nuclear (N) components of the NFAT complex⁴, the Genome Data Base Nomenclature Committee has designated the family of cytoplasmic components: NFATC. Within this family, NFATC1 (genome database number G00-407-615) was previously named NF-ATc; NFATC2 (G00-407-616) was formerly NF-ATp; NFATC3 (G00-407-617) was NF-AT4 or NF-ATx; and NFATC4 (G00-407-618) was NF-AT3.) As suggested by our earlier biochemical fractionation studies⁴, and as demonstrated for NF-ATp (ref. 9), each NF-AT family member undergoes nuclear translocation within 5 min of cellular stimulation with ionomycin (Fig. 1e, f and data not shown), or in response to the presentation of specific peptide in the context of appropriate major histocompatibility complex (MHC) molecules (Fig. 1a, b). Remarkably, curtailment of calcium signalling by the addition of FK506 or cyclosporin A (inhibitors of calcineurin²), by the removal of extracellular calcium with EGTA, or by the removal of peripheral T cells from anti-CD3 coated plates, induces rapid nuclear export of each family member (Fig. 1c, d, g, h, and data not shown for NF-ATc). Export is essentially complete within 5 min (data not shown), and occurs in the presence of 40 μM anisomycin (which blocks 98% of protein synthesis), indicating that termination of the calcium stimulus results in relocalization of pre-existing NF-AT molecules from nucleus to cytoplasm. Similar results have been obtained by western blots of fractionated cellular extracts (data not shown).

To further explore the roles of transient and sustained Ca^{2+} signals we used T-cell lines with mutations that prevent capacitative calcium entry^{17,18}. Upon stimulation, these cell lines show a transient rise in $[\text{Ca}^{2+}]_i$ that rapidly returns to baseline¹⁸. As previously shown by biochemical fractionation, this transient calcium flux is not sufficient for nuclear import of NF-AT¹⁷ (compare Fig. 2a, b with d, e). However, NF-AT will translocate if $[\text{Ca}^{2+}]_i$ is artificially increased and sustained by increasing the

FIG. 1 NF-AT family members are rapidly exported from the nucleus upon interruption of calcium signalling. Top, the Th1 haemoglobin-specific T-cell clone PL.17 stained for NF-ATp localization: a, before stimulation; b, after 20-min exposure to specific peptide (Hb β 64-76) on I-E* expressing antigen-presenting cells; c, as in b, followed by addition of 10 ng ml⁻¹ FK506 and 40 μM anisomycin (to block protein synthesis) for 15 min; d, as in c, substituting 2.5 mM EGTA for FK506. Bottom, freshly isolated thymocytes stained for NF-ATc3 localization after: e, no stimulation; f, 20-min culture with 10 ng ml⁻¹ PMA and 1 μM ionomycin; g, as in f, followed by 10 ng ml⁻¹ FK506 and 40 μM anisomycin for 5 min; or h, FK506 and anisomycin for 15 min.

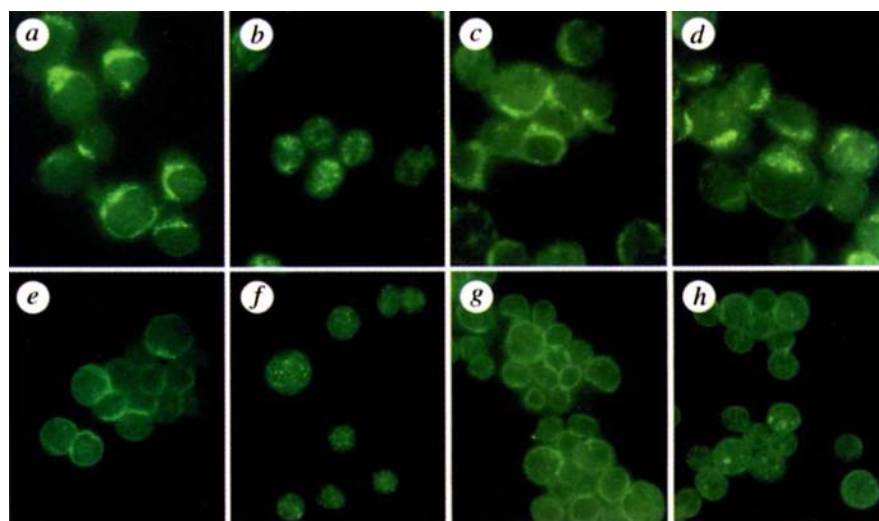
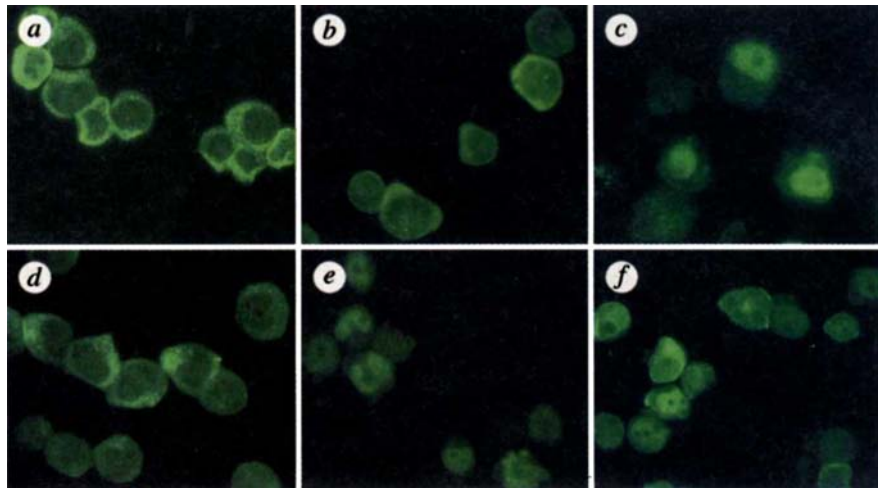


FIG. 2 NF-ATc does not translocate in a cell line with a defect in capacitative calcium entry and only transient increases in intracellular Ca^{2+} . NF-ATc localization was assessed in CJ1.1 cells¹⁸ (a–c) or the parental line Jurkat (d–f). Cells were either: a, d, cultured in media alone; b, e, stimulated with ionomycin ($1\ \mu\text{M}$) and PMA ($10\ \text{ng ml}^{-1}$) for 4 h; or c, f, stimulated as in b, but in the presence of $10\ \text{mM CaCl}_2$.



concentration of extracellular Ca^{2+} (Fig. 2c, f). Thus sustained levels of elevated intracellular calcium are required to maintain NF-AT in the nucleus.

Although the bulk of nuclear NF-AT rapidly relocalizes to the cytoplasm in normal T cells when calcium signalling is interrupted (Fig. 1), ongoing gene transcription should only be terminated if active, DNA-bound NF-AT complexes rapidly dissociate from target sequences. In agreement with previous results¹⁹, we find that the half-time for dissociation of NF-AT complexes from the strong distal human IL-2 promoter NF-AT site *in vitro* is ~ 15 min (data not shown). Because activators are required for the reinitiation of transcription *in vivo*²⁰, transcription should rapidly cease once NF-AT is exported to the cytoplasm. As predicted, properly initiated transcription measured by ribonuclease protection of either the endogenous IL-2 gene or a transfected NF-AT reporter construct is virtually undetectable 60 min after the addition of cyclosporin A to activated cells (Fig. 3, lane 6). Thus cyclosporin A treatment results in rapid nuclear export of NF-AT and termination of IL-2 gene transcription. Conversely, constitutive nuclear expression of NF-AT family members readily induces calcium-independent (Fig. 4a) and cyclosporin A-resistant (Fig. 4b) activation of an IL-2 promoter–luciferase reporter (IL-2-Luc), whereas overexpression and nuclear localization of other transcription factors implicated in IL-2 gene regulation yields minimal activity. Thus nuclear localization of NF-AT family members by maintenance of elevated intracellular calcium during early T-cell activation is required for IL-2 gene induction in Jurkat T lymphocytes.

In previous studies, several transcription factors and kinases, including NF- κ B, AP1 and JNK, have been implicated in IL-2 gene regulation and as targets of cyclosporin A and calcineurin^{21,22}. We tested the role of these calcium-regulated transcription factors by examining their ability to regulate the endogenous IL-2 gene. Although endogenous genes are chromatinized and rarely activated by a transiently transfected transcription factor²³ (reviewed in refs 24, 25), approximately 5% to 6% of cells transiently expressing nuclear NF-ATc and NF-ATp produce IL-2 after stimulation in the presence of FK506 (Fig. 5). In contrast, control transfectants expressing either HNF-1 α (an endodermal transcription factor²⁶; data not shown), NF- κ B family members p50 and p65, or AP1 family members c-Fos and JunD, or Fra-1 and JunD, demonstrate only background staining for IL-2 production. IL-2 secretion from stimulated, FK506-treated cells expressing nuclear NF-ATc and NF-ATp was confirmed by enzyme-linked immunosorbent assay (ELISA). Thus lymphocytes can be rendered resistant to FK506 simply by artificial nuclear localization of NF-ATc and NF-ATp. This indicates that an important regulatory effect of cyclosporin A and FK506 is the ability to block the nuclear localization of NF-AT family members, and suggests that the mechanistic steps resulting in

nuclear localization of NF-AT will be targets for the development of new immunosuppressants.

As NF-AT family members are expressed in most tissues^{13–15,27}, our data also provide a mechanism by which the duration of $[\text{Ca}^{2+}]_i$ changes can be translated into distinct biological responses. Calcium signals regulate both rapid, cytoplasmic responses (such as adhesion, locomotion, vesicle fusion, muscle contraction and synaptic transmission) induced by transient increases in calcium, and also delayed responses (such as proliferation and differentiation) induced by sustained increases in calcium that require transcription. We have shown that a sustained, high $[\text{Ca}^{2+}]_i$ is required to maintain NF-AT nuclear localization; this provides not only a mechanism for buffering nuclear transcriptional events from brief Ca^{2+} transients, but also a way of dynamically regulating gene expression in response to growth factors and changing environmental conditions. □

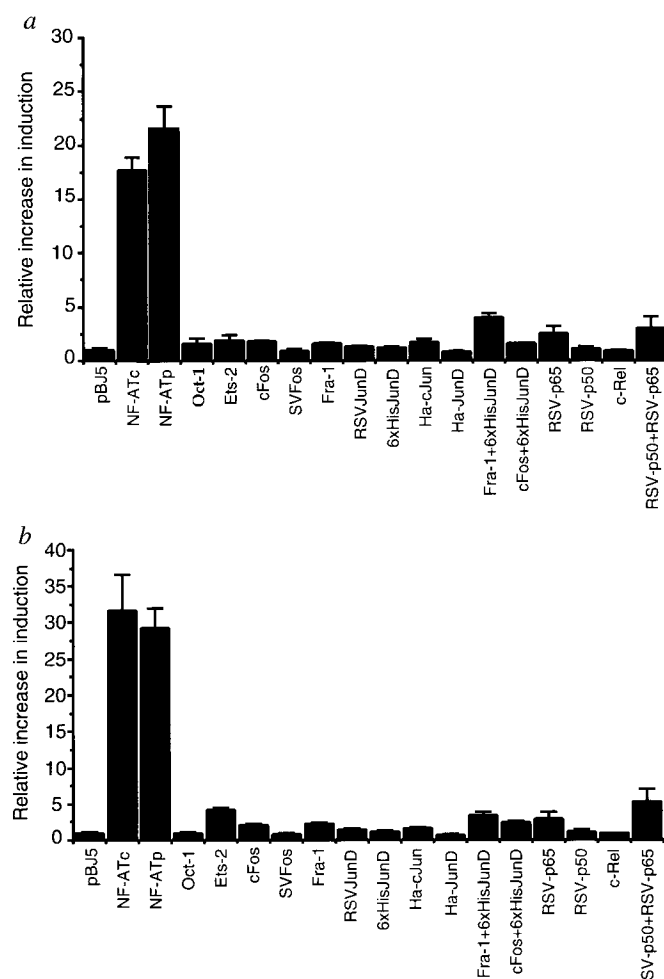
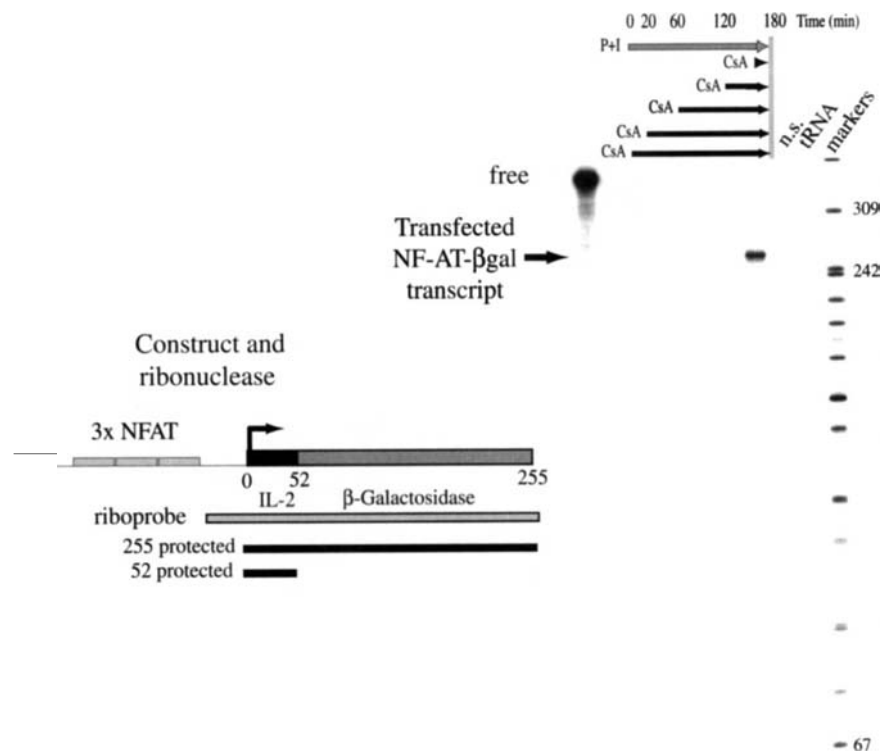
Methods

Cell culture. Stimulation and culture conditions for PL17 cells were as described²⁸. The fibroblast APC line DCEK Hi7 (stably transfected with I-E^c) was incubated overnight at 10^6 cells per well of a 24-well plate, excess and mitotic cells washed off, and adherent cells pulsed with $10\text{-}\mu\text{M}$ peptide in media, or media alone, for 2 h. The Hb β 64–76 reactive PL17 T-cell clone was used for analysis two weeks after expansion on feeder cells, after purification over a ficoll gradient (Cedarlane). PL17 cells were plated on the fibroblast line Hi-7 with or without peptide and/or FK506 ($10\ \text{ng ml}^{-1}$), anisomycin ($40\ \mu\text{M}$) and EGTA ($2.5\ \mu\text{M}$), for the times indicated in Fig. 1. The cell lines Jurkat, T-antigen Jurkat, and CJ1.1 were cultured in RPMI 1640/10% FCS. Stimuli and inhibitors included: $1\ \mu\text{M}$ ionomycin, $10\ \text{ng ml}^{-1}$ PMA, $10\ \text{mM CaCl}_2$, $40\ \mu\text{M}$ anisomycin and $2.5\ \mu\text{M}$ EGTA. Freshly isolated thymocytes were stimulated as Jurkat cells.

Immunofluorescence. All cells were centrifuged (Cytospin, Shandon, 3 min at $300\ \text{r.p.m.}$) onto poly-L-lysine coated slides ($1\ \text{mg ml}^{-1}$). PL17 cells were fixed (0.5% paraformaldehyde, Electron Microscopy Sciences, 16% solution, 10 min at room temperature), permeabilized (cold methanol for 2 min), and rehydrated (PBS for 10 min). Thymocytes were fixed and permeabilized with acetone for 15 s at -20°C and methanol for 1.5 min at -20°C . PL17 cells were stained with the NF-ATp-specific monoclonal antibody G1-D10 ($1:300$ in PBS/1% normal rat serum), followed by anti-mouse biotin ($1:2,000$, Caltag, rat adsorbed), and avidin D-FITC ($1:1,000$, Vector). Thymocytes were stained with an affinity-purified rabbit polyclonal anti NF-ATx/c3/4 ($1:200$ in 1% normal rat serum), followed by anti-rabbit biotin ($1:2,000$, Jackson), and avidin-FITC ($1:1,000$, Vector). Jurkat, T-antigen Jurkat and CJ1.1 were fixed in 4% paraformaldehyde for 10 min at room temperature, permeabilized for 2 min in methanol at room temperature, and rehydrated in PBS. The NF-ATc-specific monoclonal 7A6 (ref. 13) was used at $1:1,000$ in 1% BSA, followed by anti-mouse biotin ($1:1,000$, Vector), and avidin D-FITC ($1:1,000$, Vector).

RNAse protection. T-antigen Jurkats were stimulated with PMA ($20\ \text{ng ml}^{-1}$) and ionomycin ($1\ \mu\text{M}$) at time zero, split into equal portions, and cultured at 37°C with cyclosporin A ($100\ \text{ng ml}^{-1}$) addition at the times indicated in Fig. 3. Cells were collected after 180 min; RNA was prepared and RNAse protection studies were performed as described²⁹, using the riboprobe shown, which spans both the IL-2 minimal promoter and the β -galactosidase gene. This yields both a 255-bp protected β -galactosidase message and a 52-bp protected fragment from endogenous IL-2 transcripts.

FIG. 3 Rapid reversal of IL-2 transcription and NF-AT-dependent transcription after cyclosporin A treatment. Jurkat cells containing a single, integrated copy of three tandem repeats of the distal IL-2 NF-AT site directing transcription of β -galactosidase³⁰ were assessed for message production of IL-2 and β -galactosidase on interruption of signalling by cyclosporin A treatment. Lane 1, undigested probe; 2, blank; 3, cyclosporin A added at the time of stimulation; 4, cyclosporin A added 20 min after stimulation; 5, cyclosporin A added 60 min after stimulation; 6, cyclosporin A added 120 min after stimulation; 7, cyclosporin A added immediately before cells were collected; 8, RNA from unstimulated cells.



Endogenous IL-2 transcripts

1 2 3 4 5 6 7 8 9 10

FIG. 4 Constitutive nuclear localization of NF-ATc or NF-ATp results in calcium-independent, FK506-insensitive activation of the IL-2 promoter. T-antigen Jurkat cells (10^7) were cotransfected with expression constructs encoding a variety of transcription factors implicated in IL-2 gene regulation, and a reporter construct consisting of the complete 320-bp IL-2 promoter directing Luciferase expression. Results are expressed as average relative increase in induction over unstimulated cells, and represent results from at least three transfections. Error bars represent standard error of the mean. *a*, IL-2 Luciferase induction by PMA treatment; *b*, IL-2 Luciferase induction by PMA plus ionomycin stimulation in the presence of FK506.

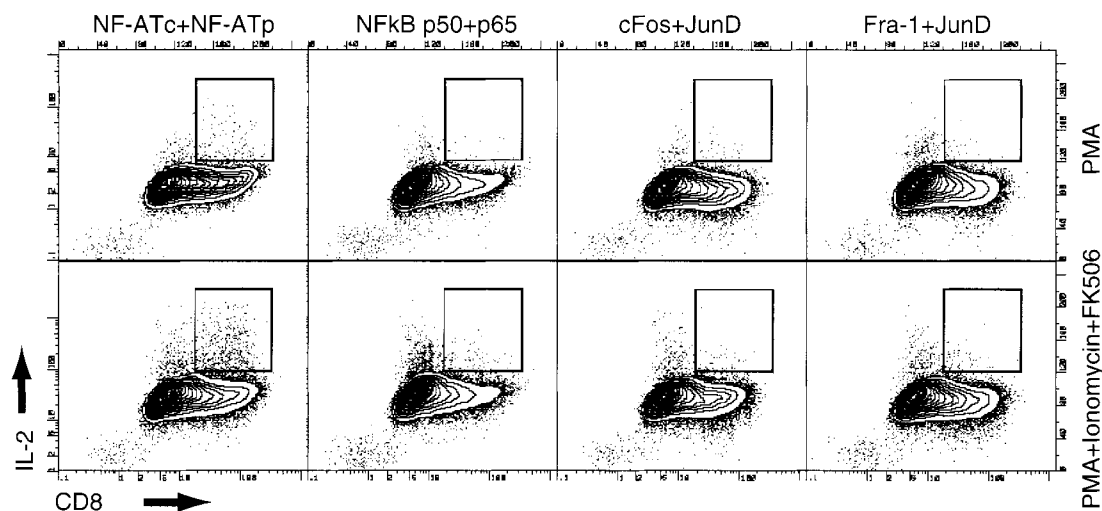


FIG. 5 Deliberate localization of NF-ATc and NF-ATp to the nucleus results in calcium-independent, FK506-insensitive activation of the IL-2 gene. Jurkat cells were cotransfected with a plasmid directing constitutive expression of CD8, and plasmids directing constitutive expression of NF-ATc and NF-ATp. The NF-AT constructs encoded molecules with small N-terminal deletions and modifications which results in NF-AT constitutive nuclear localization. Transfected cells are CD8-cyochrome positive, and IL-2 producing cells are IL-2-PE positive. The example shown is representative

of three independent experiments in which the percentage of cells transfected with NF-ATc that expressed IL-2 ranged from 1% to 2.7% when treated with PMA + FK506, and from 4.3% to 5.9% when stimulated with PMA + FK506 + ionomycin. In contrast, control transfections with HNF- α , NF-kB p65 + p50, or AP1 family members cFos + JunD or Fra-1 + JunD showed only background levels of IL-2-positive cells with these stimuli. CD8-positive transfectants given full stimulation (PMA + ionomycin) routinely were 20% to 40% IL-2 positive, depending on the length of stimulation.

Transient transfections. T-antigen Jurkats were transiently cotransfected with an IL-2-Luciferase reporter and expression constructs encoding the transcription factors detailed in Figs 4 and 5. All transcription factors were expressed under the control of the SR- α promoter except for SVFos (SV40 early promoter), RSVJunD, RSV-p65 and RSV-p50 (Rous sarcoma virus promoter). Cells were electroporated (BIO-RAD Gene Pulser; 240 V, 960 μ F, 0.4 cm gap cuvette) at room temperature in 250 μ l culture media, and cultured in 10 ml media overnight at 37 °C. Transfections were divided into equal parts and cells were either left unstimulated or were stimulated with either PMA alone (20 ng ml⁻¹), PMA and ionomycin (1 μ M), or PMA, ionomycin and FK506 (5 ng ml⁻¹) for 7–12 h before being collected. Luciferase assays were performed using standard techniques. For endogenous IL-2 production, Jurkats were transiently cotransfected with expression constructs encoding a cytoplasmic tail negative variant of CD8, SV40 large T antigen (SV-Tag), and one or two expression constructs encoding various NF-AT, NF-kB or AP1 family members. Cells were cultured overnight, live cells purified by ficoll density gradient sedimentation, and stimulated for 4–6 h as in Fig. 4. IL-2 secretion was inhibited by 2-h treatment with monensin (2 μ M), cells were stained for surface CD8 (FITC or Cyochrome, PharMingen) by standard techniques, fixed in 2% formaldehyde, permeabilized with 0.5% saponin, and stained for intracellular IL-2 with anti-IL-2-PE (PharMingen) in the presence of 0.5% saponin²⁸, allowing simultaneous detection of cell-surface CD8 and intracellular IL-2 proteins.

23. Turner, C. A. Jr, Mack, D. H. & Davis, M. M. *Cell* **77**, 297–306 (1994).
24. McKnight, S. L. & Yamamoto, K. R. *Transcriptional Regulation Vols I and II* (Cold Spring Harbor Laboratory Press, New York, 1992).
25. Weintraub, H. *et al. Science* **251**, 761–766 (1991).
26. Courtois, G., Morgan, J. G., Campbell, L. A., Fourel, G. & Crabtree, G. R. *Science* **238**, 688–692 (1987).
27. Ho, A. M., Jain, J., Rao, A. & Hogan, P. G. *J. Biol. Chem.* **269**, 28181–28186 (1994).
28. Sander, B., Hoiden, I., Andersson, U., Moller, E. & Abrams, J. S. *J. Immunol. Meth.* **166**, 201–214 (1993).
29. Melton, D. A. *et al. Nucleic Acids Res.* **12**, 7035–7056 (1984).
30. Fiering, S. *et al. Genes Dev.* **4**, 1823–1834 (1990).

ACKNOWLEDGEMENTS. We thank P. Allen for the use of the T-cell line PL17, and R. Lewis and C. Fanger for the use of the Jurkat mutant CJ1.1.

CORRESPONDENCE and requests for materials should be addressed to G.R.C. (e-mail: hf.grc@forsthe.standord.edu). NF-ATc nomenclature in the Genome Data Base is available on <http://gdbwww.gdb.org/gdb-bin/gdb/lite/bin/genes>.

A meiotic recombination checkpoint controlled by mitotic checkpoint genes

David Lydall*, Yuri Nikolsky†, Douglas K. Bishop† & Ted Weinert*

* Department of Molecular and Cellular Biology, University of Arizona, Tucson, Arizona 85721, USA

† Department of Radiation and Cellular Oncology/Department of Molecular Genetics and Cell Biology, University of Chicago, Chicago, Illinois 60638, USA

In budding yeast, meiotic recombination occurs at about 200 sites per cell and involves DNA double-strand break (DSB) intermediates^{1–3}. Here we provide evidence that a checkpoint control requiring the mitotic DNA-damage checkpoint genes *RAD17*, *RAD24* and *MEC1* ensures that meiotic recombination is complete before the first meiotic division (MI). First, *RAD17*, *RAD24* and *MEC1* are required for the meiotic arrest caused by blocking the repair of DSBs with a mutation in the *recA* homologue *DMC1*.

Received 12 July; accepted 22 August 1996.

1. Klee, C. B., Crouch, T. H. & Krinks, M. H. *Proc. Natl Acad. Sci. USA* **76**, 6270–6273 (1979).
2. Liu, J. *et al. Cell* **66**, 807–815 (1991).
3. Clipstone, N. A. & Crabtree, G. R. *Nature* **357**, 695–697 (1992).
4. Flanagan, W. F., Cothesy, B., Bram, R. J. & Crabtree, G. R. *Nature* **352**, 803–807 (1991).
5. Nowell, P. C. *Cancer Res.* **20**, 462–466 (1960).
6. Whitney, R. B. & Sutherland, R. M. *J. Cell Physiol.* **80**, 329–337 (1972).
7. Tsien, R. Y., Rozan, T. & Rink, T. *Nature* **295**, 68–71 (1982).
8. Weiss, A., Shields, R., Newton, M., Manger, B. & Imboden, J. *J. Immunol.* **138**, 2169–2176 (1987).
9. Shaw, K. T.-Y. *et al. Proc. Natl Acad. Sci. USA* **92**, 11205–11209 (1995).
10. Rao, A. *Immunol. Today* **15**, 274–281 (1994).
11. Crabtree, G. R. & Clipstone, N. A. *Annu. Rev. Biochem.* **63**, 1045–1083 (1994).
12. Hodge, M. R. *et al. Immunity* **4**, 397–405 (1996).
13. Northrop, J. P. *et al. Nature* **369**, 497–502 (1994).
14. Ho, S. N. *et al. J. Biol. Chem.* **270**, 19898–19900 (1995).
15. Hoey, T., Sun, Y.-L., Williamson, K. & Xu, X. *Immunity* **2**, 461–472 (1995).
16. Masuda, E. S., Naito, Y., Tokumitsu, H. *et al. Mol. Cell Biol.* **15**, 2697–2706 (1995).
17. Serafini, A. T., Lewis, R. S., Clipstone, N. A. *et al. Immunity* **3**, 239–250 (1995).
18. Fanger, C. M., Hoth, M., Crabtree, G. R. & Lewis, R. S. *J. Cell Biol.* **131**, 655–667 (1995).
19. Jain, J., Miner, Z. & Rao, A. *J. Immunol.* **151**, 837–848 (1993).
20. Ho, S. N., Biggar, S. R., Spencer, D. M., Schreiber, S. L. & Crabtree, G. R. *Nature* **382**, 822–826 (1996).
21. Emmel, E. A. *et al. Science* **246**, 1617–1620 (1989).
22. Su, B. *et al. Cell* **77**, 727–736 (1994).

Second, *mec1* and *rad24* single mutants (*DMC1*⁺) appear to undergo MI before all recombination events are complete. Curiously, the mitosis-specific checkpoint gene *RAD9* is not required for meiotic arrest of *dmc1* mutants⁴. This shows that although mitotic and meiotic control mechanisms are related, they differ significantly. Rad17 and Rad24 proteins may contribute directly to formation of an arrest signal by association with single-strand DNA in mitosis and meiosis.

In mitotic cells, DSBs and other types of DNA damage cause checkpoint-control-mediated cell-cycle arrest⁵⁻⁷. The checkpoint genes *RAD9*, *RAD17* and *RAD24* encode proteins that may interact directly with DNA during repair⁸; *MEC1* encodes a protein that may act in a signal transduction pathway that links DNA damage with cell-cycle arrest⁹⁻¹¹. To determine whether any mitotic checkpoint genes are required for the meiotic arrest caused by *dmc1* mutations (*dmc1* arrest), we constructed checkpoint *dmc1* double mutants and induced them to undergo meiosis. The kinetics of chromosome segregation were assessed by staining cells with the DNA-specific stain DAPI (Fig. 1). The *dmc1* single mutants and *rad9 dmc1* double mutants arrested in prophase, as previously shown⁴. However, other checkpoint *dmc1* double mutants proceeded through two meiotic divisions with essentially normal kinetics (Fig. 1 and Table 1). Therefore, *dmc1* arrest required the mitotic checkpoint genes *RAD17*, *RAD24* and *MEC1*. The requirement for checkpoint genes in meiotic arrest was not a specific feature of the *dmc1* defect because the meiotic delay caused by another mutation, *zip1* (refs 12, 13), also depended on *RAD24* (Fig. 1). The failure of checkpoint *dmc1* double mutants to arrest resulted in the formation of aberrant meiotic products (Table 1).

The observation that mitotic DNA-damage checkpoint genes were required for *dmc1* arrest suggested that the checkpoint genes are required to promote arrest when meiotic recombination intermediates are present. However, there were other explanations for the alleviation of *dmc1* arrest. First, mutations that block the recombination pathway at early stages also prevent *dmc1* arrest: such mutations include *spo11* and *rad50* deletions, which prevent DSB formation^{2,14-16}, and the *rad50S* mutation, which prevents the formation of single-stranded DNA (ssDNA) that normally results from resection at DSBs^{4,17}. Second, it was possible that the checkpoint mutations alleviated arrest by allowing resolution of recombination intermediates by a *DMC1*-independent pathway^{18,19} (Y.N. and D.K.B. unpublished results). We now describe how we eliminated these alternatives.

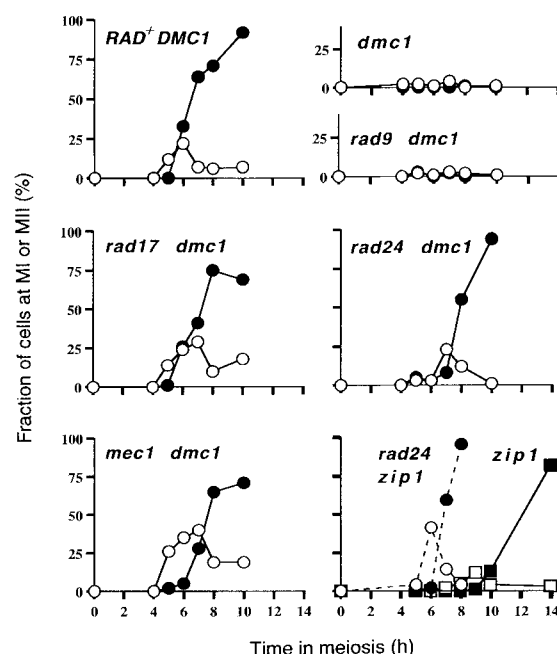


FIG. 1 Mitotic checkpoint genes are required for meiotic arrest of *dmc1* mutants. Diploid yeast cells were induced to enter a synchronous meiotic cycle. At various times, cells were removed, fixed in 70% ethanol, and the meiotic stage determined by staining nuclear DNA with 4,6-diamino-2-phenylindole (DAPI). Cells with two nuclei had completed MI (open circles); cells with >2 nuclei (or DNA-staining spots) had completed MII (filled circles); *zip1* single mutants (squares). At least 100 cells were scored by fluorescence microscopy for each time point. Strains: wild type, DKB282; *dmc1*, DKB217; *rad17 dmc1*, DDY22; *rad24 dmc1*, TY11; *mec1 smX dmc1*, DDY17; *rad9 dmc1*, DDY28; *zip1*, TY62; *zip1 dmc1*, TY65.

The effect of checkpoint mutations on DSB metabolism at the previously characterized recombination hotspot *his4::LEU2* was determined by Southern analysis^{2,4,17}. DNA was purified from cells at various times after induction of meiosis and tested for the presence of DSBs at *his4::LEU2* by indirect end-labelling (Fig. 2). We used a single-stranded RNA probe that specifically recognizes the 3' strands of DNA flanking DSBs (to avoid the complications

TABLE 1 Role of mitotic checkpoint genes in meiosis

Genotype	Percentage of cells post-MI at 10 h	Percentage of cells forming tetrads	Spore viability (%)	Meiotic recombination frequency (%)		Number of Rad51-containing foci (average per nucleoid)		Strains
				at <i>HIS4</i>	at <i>ARG4</i>	Pre-MI	Post-MI	
Wild type	95	77.2 ± 11.9	93.2 ± 9.9	0.34 ± 0.12	1.73 ± 0.66	35.7 ± 10.9	<1 in 500	TY19, DKB282
<i>rad9Δ</i>	89	83	90	0.66	1.4			DDY30
<i>mec1-1</i>	86	62	51.3 ± 19.6	0.29 ± 0.18	1.75 ± 0.28	30.7 ± 9.2	1.5 ± 0.6	DDY15, DKB836
<i>rad17Δ</i>	92	87	18.3 ± 5.1	0.31	1.76			DDY9
<i>rad24Δ</i>	89	55 ± 6.5	15.3 ± 5.9	0.67	1.4 ± 0.8	30.4 ± 11.2	2.3 ± 1.1	TY11, 12 DDY20
<i>dmc1Δ</i>	1	<1	1.2			28.4 ± 8.6	NA	TY17, DKB217
<i>dmc1::LUK10 rad9Δ</i>	2	<1	0.125					DDY28
<i>dmc1Δ mec1-1</i>	89	<1	<0.1			22.7 ± 7.2	20.8 ± 7.5	DDY17, DKB839
<i>dmc1Δ rad17Δ</i>	91	<1	<0.1					DDY22
<i>dmc1Δ rad24Δ</i>	87	<1	<0.1			25.1 ± 10.2	34.9 ± 11.2	TY15, DDY20
<i>zip1Δ</i>	17	34 ± 2.0	46.4 ± 9.1					TY62
<i>zip1Δ rad24Δ</i>	93	18.5 ± 3.0	3.0 ± 0.0					TY65

The fraction of cells post-MI at 10 h was determined by DAPI staining as for Fig. 1. The fraction of cells forming tetrads (scored by bright-phase microscopy) was determined after 24 h in sporulation medium. Spore viability was determined after random spore analysis or from dissected tetrads. Meiotic recombination between *arg4-BglII* and *arg4-Nsp1* or between *his4X* and *his4B* heteroalleles was determined by plating spores (prepared for random spore analysis) on plates lacking arginine or histidine, and is calculated as a fraction of the viable spores. For Rad51 focus analysis, Pre-MI and post-MI nuclei were identified based on associated tubulin staining structures (see Fig. 3 legend and Methods).

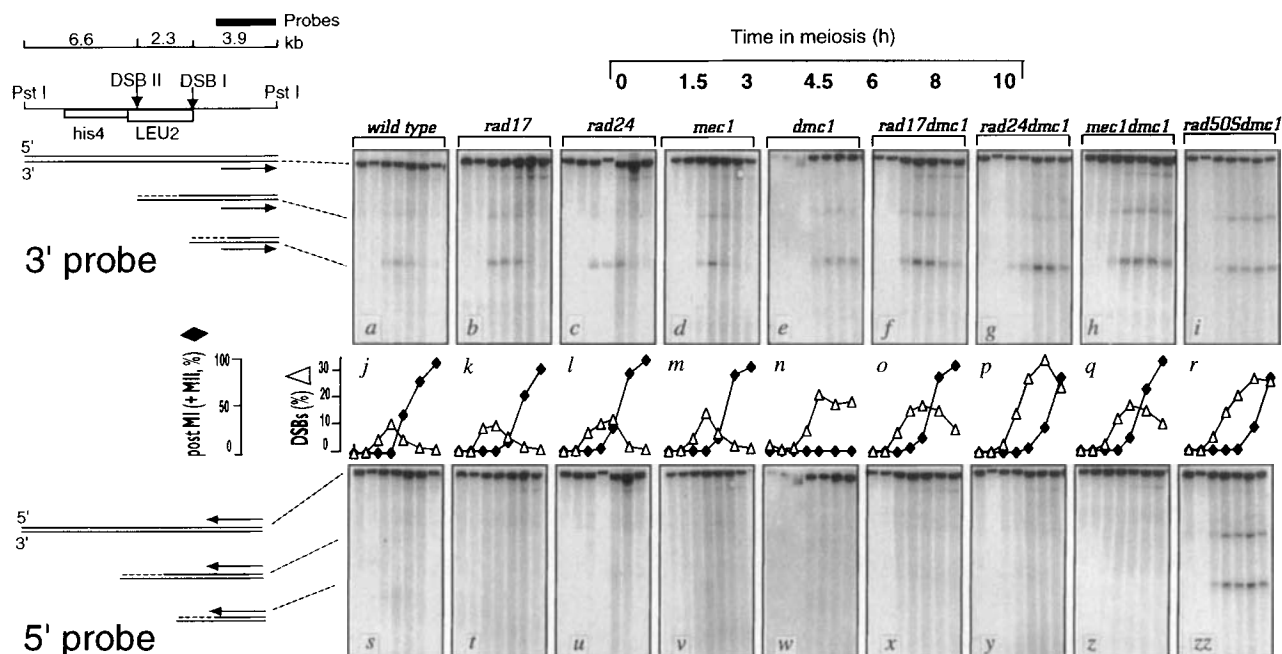


FIG. 2 Checkpoint genes are not required for the formation or resolution of DSBs or for 5' resection at DSBs. Left, representation of the *his4::LEU2* locus, which contains two prominent DSBs induced during meiosis. These DSBs yield fragments of 3.9 kb and 6.2 kb when DNA is cut with *Pst*I. Right, kinetics of DSB appearance as cells progress through meiosis. To identify DSBs, DNA was cut with *Pst*I, separated on alkaline agarose gels, trans-

ferred to nylon membranes and hybridized with strand-specific ssRNA probes⁴. Gels a–i show analyses of 3' strands; j–r, plots of the amount of DNA present at DSB1 (3.9-kb band) as a fraction of the parental (12.8-kb band) (triangles), and the fraction of cells that had completed MI (including cells that completed both MI and MII) is shown (diamonds), determined as for Fig. 1. Gels s–zz show analyses of the 5' strands of DNA.

of analysing double-stranded DNA caused by variable and extensive resection of 5' strands^{2,4}). All three *checkpoint dmc1* double mutants were similar to the *dmc1* single mutant; DSBs formed normally and most were retained for at least 10 hours (Fig. 2e–i, n–r). By 10 hours, most *checkpoint dmc1* mutant cells had executed MI (Fig. 2n–r). We conclude that *checkpoint dmc1* mutants produce normal numbers of DSBs and that they enter MI in the presence of unresolved DSBs. Consistent with this interpretation, *checkpoint dmc1* double mutants produced meiotic products containing fragmented DNA (as determined by DAPI staining) and very few spores (data not shown). Furthermore, 99.9% of spores produced were dead (Table 1).

The effect of the checkpoint mutations on formation of ssDNA by 5' resection of DSB ends was determined using a probe specific for the 5' strand at DSB sites. If resection occurs normally, 5'-strand DSB fragments are essentially undetectable by our method. If 5' strands are not degraded, as in *rad50S* strains, DSB fragments are readily detected (Fig. 2zz). 5'-strand DSB fragments were not detected in any of the *checkpoint dmc1* mutants (Fig. 2w–z). We conclude that the alleviation of *dmc1* arrest caused by checkpoint mutations is not due to the absence of ssDNA at DSB sites.

We used a sensitive assay that detects protein complexes associated with recombination intermediates to provide additional evidence that *checkpoint dmc1* mutants enter MI in the presence of unresolved recombination intermediates. Rad51 protein, a homologue of the Dmc1 protein²⁰, forms multiple complexes during meiotic prophase²¹. These complexes are detected as subnuclear foci by fluorescent immunostaining of spread meiotic nuclei. Rad51 foci are likely to mark the sites of DSBs because their appearance and disappearance is associated with DSBs in wild-type and mutant cells²¹ (Y.N. and D.B., unpublished data). Anti-tubulin staining of spindles was used to distinguish cells that had initiated MI from those that had not. We found that all strains had similar numbers of Rad51 foci before formation of the MI spindle (up to 36 complexes per nucleus) (Fig. 3 and Table 1); again, this shows that the checkpoint mutations do not block

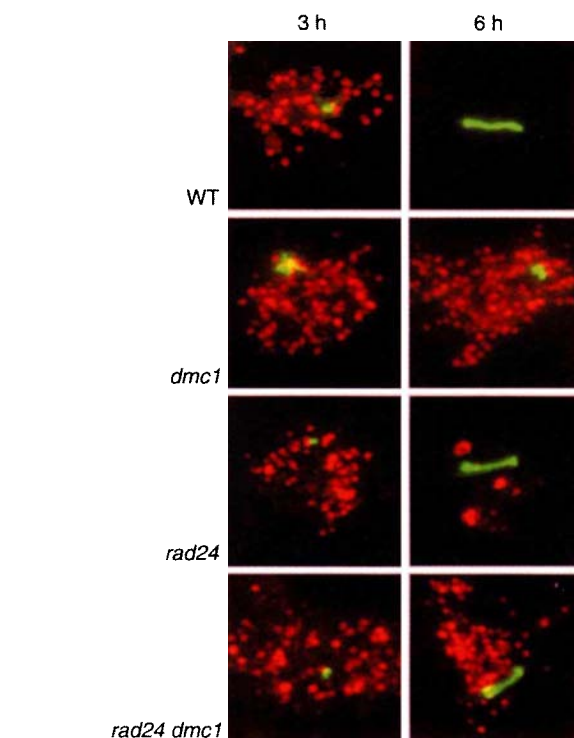


FIG. 3 Analysis of Rad51 foci in cells undergoing meiosis suggests that recombination intermediates persist in checkpoint mutants entering MI. Spread meiotic nuclei stained with anti-Rad51 (red) and anti-tubulin (green) antibodies. Representative nuclei are shown for each strain. Left panels, examples of spread nuclei with near-maximal numbers of Rad51 foci and a single tubulin focus (microtubule-organizing centre), indicating that spindle pole bodies are unduplicated or unseparated. Right panels, nuclei with elongated tubulin structures, as seen when spindle pole bodies separate to form the MI spindle.

meiosis before the Dmc1-dependent stage. We next examined cells that had entered MI to determine whether they retained Rad51 foci. Rad51 foci were not detected in MI spindle-associated wild-type nuclei (<1 focus per 500 cells). This result was expected, because recombination is normally complete before MI. In contrast, MI spindle-associated nuclei from *checkpoint dmc1* double mutants contained multiple Rad51 foci, providing evidence that these cells enter MI without completing recombination.

This analysis showed that a checkpoint control is required for arrest when completion of recombination is blocked by a *dmc1* mutation. The following analysis indicates that checkpoint controls ensure that recombination is complete before MI even in normal cells. Characterization of checkpoint single mutants revealed no role for checkpoint genes in DSB formation or resolution (Fig. 2a–d, j–m). Intragenic recombination among surviving meiotic products (spores) was also normal (Table 1). However, when we used the sensitive method of Rad51–tubulin double staining, MI nuclei from checkpoint single mutant (*DMC1*⁺) cells were typically found to contain one or two Rad51 foci, suggesting that a small fraction of recombination events were incomplete at the time of MI (Fig. 3 and Table 1). Failure to complete one or a few recombination events could explain the inviability of 50–85% of spores formed by *checkpoint* single mutants (Table 1)⁸. We favour the possibility that a checkpoint control is activated during normal meiosis, allowing time for the completion of a few slow recombination events. However, we cannot exclude other explanations: for example, *RAD24* and *MEC1* may be required directly for recombination and/or repair at a few sites. *MEI41*, a *Drosophila* homologue of *MEC1* (ref. 22), is required for the normal frequency and distribution of meiotic crossover recombination events, as well as for mitotic checkpoint control²³. Whether *MEC1* has a role in meiotic recombination like that of *MEI41* is unclear (Table 1; but see also Methods and ref. 24).

We conclude that related checkpoint mechanisms monitor both meiotic recombination and mitotic DNA damage because both are dependent on *RAD17*, *RAD24* and *MEC1*. We suggest that a nucleoprotein complex, whose components include ssDNA, Rad17 and Rad24, is required to generate the primary signal for arrest in both meiosis and mitosis. In meiosis, ssDNA is present in each of the four recombination mutants that arrest after DSBs are formed (*dmc1*, *rad51*, *rad52* and *zip1*; refs 4, 19, 25, and L. Xu, personal communication) but is absent in *rad50S* mutants that also make DSBs but do not arrest^{4,17}. In mitosis, ssDNA is associated with arrest caused by defective DNA replication (near telomeres in *cdc13* mutants)²⁶ and with that caused by DSBs^{5,27}. The evidence that checkpoint proteins bind DNA comes primarily from studies on mitotic checkpoint control. In mitotic cells, Rad17 and Rad24 appear to be involved in both DNA degradation and checkpoint control in *cdc13* mutants⁸. Furthermore, *RAD17* encodes a putative 3'–5' DNA exonuclease⁸. Thus, although we did not detect a role for *RAD17* and *RAD24* in DNA degradation at meiotic DSBs, it is likely that the products of these genes also associated with ssDNA at the meiotic recombination checkpoint. Therefore in meiosis, and perhaps also in mitosis, Rad17 and Rad24 may need only associate with ssDNA to promote an arrest signal. Although it shares key features with mitotic checkpoint control, the mechanism that monitors meiotic recombination appears to be distinct. *RAD9* is required for arrest of mitotic and early meiotic cells with damage²⁸, but not for arrest at the later recombination checkpoint (see refs 8, 29 for further discussion). There is also recent evidence that meiosis-specific recombination genes play a role in checkpoint control¹⁹. □

Methods

Strains. All strains were in the SK1 background and generally had the genotype *MATa/MATα leu2::hisg⁺ura3⁺ho::LYS2⁺lys2⁺arg4BglII/arg4Nsp1 his4X::*

LEU2::URA3/his4B::LEU2. The *rad9*, *rad17*, *rad24* and *zip1* disruptions were introduced by transformation (genotypes of specific strains available upon request). The *mec1-1* mutation was backcrossed four times into the SK1 background from the A364a background. The *mec1-1* mutation is probably a null mutation kept viable by an unlinked suppressor, *sm1*, which suppresses the essential function of *MEC1*, but not its checkpoint functions (A. Paulovich and R. Rothstein, personal communications). We have not rigorously shown that *sm1* is the suppressor in SK1 outcrosses (hence *sm1X*), but we have verified that the loss of checkpoint control in *mec1-1 sm1X dmc1* mutants is due to the *mec1-1* mutations because *dmc1* arrest was restored when *MEC1* was reintroduced by transformation (data not shown). An earlier study found that some alleles of *MEC1* (*ESR1*) reduced recombination frequency when meiotic cells were returned to mitotic growth during recombination²⁴, but it is not clear whether this defect indicates that Mec1 has a direct role in recombination or in mitotic checkpoint control.

Analysis of DSBs. Diploid yeast cells were induced to enter a synchronous meiotic cycle at 30 °C using standard methods^{4,17}. At appropriate times, 15-ml aliquots of cells were fixed by adding ethanol to a final concentration of 50%. DNA was isolated after storage overnight at 4 °C. A standard DNA isolation procedure was used², except that spheroplasting and proteinase K treatments were done in the presence of 100 mM EDTA (zymolyase was in 1 M sorbitol, 100 mM EDTA, 10 mM sodium phosphate, pH 7.2; 20 µg ml⁻¹ proteinase K was in 100 mM NaCl, 100 mM EDTA, 50 mM Tris–HCl, pH 7.5, 0.3% SDS). After digesting the DNA with *Pst*I, it was precipitated and separated on alkaline agarose gels^{4,30} which were run for 18 h at 4 °C with buffer circulation at 45 V. Gels were then washed for 10 min in water, 2 × 10 min in 0.25 M HCl (to ensure even transfer of the 12.8-kb parental band), and again in water before being transferred to Zetaprobe nylon membrane in 0.4 M NaOH. After hybridization with ssRNA probes, radioactivity was quantified using a phosphorimager. Virgin blots were used for 3' analysis. To analyse 5' strands, blots that had been hybridized with 3' probes were stripped (0.4 M NaOH for 2 h) and reprobed with 5' probes (results were indistinguishable from virgin blots). Strains were DKB282, wild type; DDY9, *rad17::hisg*; DDY20, *rad24::LEU2*; DDY15, *mec1-1 sm1X*; DKB217, *dmc1::ARG4*; DDY22, *rad17::hisG dmc1::ARG4*; DDY18, *rad24::URA3 dmc1::LEU2*; DDY17, *mec1-1 sm1X dmc1::ARG4*; DKB170, *rad50S dmc1::LUK10*.

Immunocytology. Spread nuclei were prepared from synchronized meiotic cells as described²¹, except that lipso was omitted to preserve microtubules. Samples were double-stained with a rabbit polyclonal anti-Rad51 serum (gift from A. Shinohara) at 1 in 200 dilution and a rat monoclonal antibody against β-tubulin (YOL1/34; Accurate Chemicals) at a 1 in 1,000 dilution. Microscopy and imaging have been described²¹.

Received 3 May; accepted 3 September 1996.

1. Sun, H., Treco, D., Schultes, N. P. & Szostak, J. W. *Nature* **338**, 87–90 (1989).
2. Cao, L., Alani, E. & Kleckner, N. *Cell* **61**, 1089–1101 (1990).
3. Wu, T.-C. & Lichten, M. *Science* **263**, 515–518 (1994).
4. Bishop, D. K., Park, D., Xu, L. & Kleckner, N. *Cell* **69**, 439–456 (1992).
5. Sandell, L. L. & Zakian, V. A. *Cell* **75**, 729–739 (1993).
6. Hartwell, L. H. & Weinert, T. A. *Science* **246**, 629–634 (1989).
7. Lydall, D. & Weinert, T. *Curr. Opin. Gen. Dev.* **6**, 4–11 (1996).
8. Lydall, D. & Weinert, T. *Science* **270**, 1488–1491 (1995).
9. Weinert, T. A., Kiser, G. L. & Hartwell, L. H. *Genes Dev.* **8**, 652–665 (1994).
10. Sanchez, Y. et al. *Science* **271**, 357–360 (1996).
11. Sun, Z., Fay, D. S., Marini, F., Foliani, M. & Stern, D. F. *Genes Dev.* **10**, 395–406 (1996).
12. Sym, M., Engebrecht, J. & Roeder, G. S. *Cell* **72**, 365–378 (1993).
13. Sym, M. & Roeder, G. S. *Cell* **79**, 283–292 (1994).
14. Game, J. C., Jamb, T. J., Braun, R. J., Resnick, M. & Roth, R. M. *Genetics* **94**, 51–68 (1980).
15. Malone, R. E. & Esposito, R. E. *Mol. Cell. Biol.* **1**, 891–901 (1981).
16. Klapholz, S., Waddell, C. S. & Esposito, R. E. *Genetics* **110**, 187–216 (1985).
17. Alani, E., Padmore, R. & Kleckner, N. *Cell* **61**, 419–436 (1990).
18. Schwacha, A. & Kleckner, N. *Cell* **76**, 51–63 (1994).
19. Xu, L. & Kleckner, N. *Genes Dev.* (in the press).
20. Shinohara, A., Ogawa, H. & Ogawa, T. *Cell* **69**, 457–470 (1992).
21. Bishop, D. K. *Cell* **79**, 1081–1092 (1994).
22. Hari, L. K. et al. *Cell* **82**, 815–821 (1995).
23. Baker, B. & Carpenter, A. T. C. *Genetics* **71**, 255–286 (1972).
24. Kato, R. & Ogawa, H. *Nucleic Acids Res.* **22**, 3104–3112 (1994).
25. Ogawa, T. et al. *Cold Spring Harb. Symp. Quant. Biol.* **58**, 567–576 (1993).
26. Garvik, B., Carson, M. & Hartwell, L. *Mol. Cell. Biol.* **15**, 6128–6138 (1995).
27. White, C. I. & Haber, J. E. *EMBO J.* **9**, 663–673 (1990).
28. Weber, L. & Byers, B. *Genetics* **131**, 55–63 (1992).
29. Weinert, T. A. & Lydall, D. in *DNA Damage and Repair* (eds Nickoloff, J. A. & Hoekstra, M.) (Humana in the press).
30. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, NY, 1989).

ACKNOWLEDGEMENTS. We thank S. Roeder for strains, N. Kleckner for discussions and for sharing unpublished results, and A. Admire for help with spore viability measurements. This work was supported by grants from NIGMS to T.W. and D.K.B.

CORRESPONDENCE and requests for materials should be addressed to T.W. (e-mail: TedWeinert@tikal.biosci.arizona.edu).

Guide to authors of contributions to *Nature*

Nature is a weekly international journal covering all the sciences. It is intended for an interdisciplinary readership, so all manuscripts should be written clearly and simply. Contributors should pay particular attention to the fact that English is not the first language of many readers. Space is limited and competition severe, so brevity is highly valued. The length limits below are specified in terms of the number of pages taken up in *Nature*. An uninterrupted page of text contains about 1,300 words, so a typical Letter to *Nature*, occupying 2½ pages, can contain about 1,500 words of text and four modest display items (figures and/or tables) with brief legends. There are no formal limits on the number of display items, but extra display material must be accompanied by commensurate reductions in the length of the text. Length limits can, however, be varied upwards or downwards at the editors' discretion.

Nature publishes the following types of contribution:

Review Articles survey recent developments in a topical area of scientific research. They should not generally occupy more than 6 pages of *Nature*.

Progress articles are relatively informal reviews of particularly topical and fast-moving fields. They do not exceed 4 pages of *Nature*.

Contributors wishing to submit a Review Article or Progress should first send a one-page synopsis to the Reviews Coordinator, from whom guidelines are available on request.

Articles are reports of original scientific research whose conclusions represent a substantial advance in understanding of an important problem and are of broad general interest. They do not generally exceed 5 pages of *Nature* and have no more than 50 references. They have a 'heading' of up to 80 words, often rewritten by *Nature's* editors, which advertises the content of the paper in general terms and should not contain numbers, abbreviations or measurements unless central to the message of the paper. The introduction and summary should be contained in the first three or four paragraphs of the main text. Layout of methods, figures and legends is summarized below.

Letters to *Nature* are short reports of original research focused on an outstanding finding whose implications are of interest to nonspecialists. They should not generally exceed 2½ pages of *Nature*. They have a bold first paragraph of not more than 180 words summarizing the background, rationale, conclusions and main implications of the work, aimed particularly at nonspecialist readers. Letters should have no more than 30 references. Layout of methods, figures and legends is summarized below.

Preparation of Articles and Letters

All manuscripts should be written in English, typed double-spaced on one side of the paper only. Manuscripts submitted in electronic form cannot be considered.

Manuscripts should be accompanied by a brief covering letter from the corresponding author, containing a full postal address, telephone and fax numbers, and e-mail address.

Five copies of manuscripts and original figures are needed, together with two copies of the covering letter. Five copies of manuscripts by any of the authors that have been submitted or are in press elsewhere should also be included, clearly marked as such; failure to do so may prejudice consideration of the manuscript.

Unless otherwise instructed by the editor handling a manuscript, when sending revised or resubmitted manuscripts five copies are required, each accompanied by a copy of the authors' response to referees' and editors' comments on the earlier version.

Titles should not exceed three lines of 31 characters each (including spaces), and should avoid numerical values, acronyms, abbreviations and punctuation wherever possible.

References should be numbered sequentially as they appear in the text, tables and figure legends. Only papers that have been published or submitted to a journal should be in the numbered list: manuscripts in preparation should be mentioned in the text with a list of authors. Reference lists should not contain textual material, grant details or acknowledgements.

Acknowledgements must be brief and appear after the reference list. *Nature* does not publish grant or contribution numbers.

Figures should not be larger than 22 by 28 cm unless unavoidable. They should not contain more than one panel unless the assembled parts are logically connected. Each copy should be marked with the corresponding author's name and, when known, the manuscript reference number. One photocopy of the original figures should be provided. All original figures are returned when *Nature* cannot offer to

publish a manuscript, but one copy of the manuscript and a photocopy of the figures are kept in confidential files for three months and then destroyed. Detailed instructions for figure preparation are available from the production department, including details of acceptable digital formats.

Colour figures are welcome but a contribution towards the cost of reproduction is requested. Inability to pay this charge will not prevent *Nature* publishing figures where colour is essential.

Tables. Symbols and abbreviations in a table should be defined immediately below the table, followed by essential descriptive material, all in double-spaced text. Tables should each be presented on a separate sheet of paper.

Methods. Brief details of materials and methods can be given in the text. Otherwise, they should be described at the end of the text in a section entitled "Methods", subdivided by bold headings referring to the methods used. This section should not exceed 800 words and ideally should be shorter. Additional material can be deposited on *Nature's* Web site. In contributions with methods sections, figure legends should not exceed 150 words each, and should be confined to a description of the figure and the symbols used.

Materials. As a condition of publication, authors are required to make materials and methods used freely available to academic researchers for their own use. Supporting datasets must be made available at the time of publication from the authors directly, by posting on *Nature's* Web site, by deposition in the appropriate public database or by distribution on the Internet. In the case of X-ray crystallographic coordinates only, public access may be delayed for up to one year after publication.

Articles and Letters should both contain a statement at the end of the text: "Correspondence and requests for materials to xxxx", with an e-mail address if desired, followed by the accession numbers or Web site address for supporting datasets. The designated author need not necessarily be the senior author, but *Nature* will expect the named author to handle matters arising from published contributions.

Manuscripts can be submitted to the Editor at *Nature*, Porter's South, Crinan St, London N1 9XW or at *Nature*, 968 National Press Building, Washington, DC 20045-1938. Proofs should be returned by express mail to London. Fax numbers and e-mail addresses for all editorial offices are on *Nature's* masthead.

Other contributions

Nature publishes informal contributed material as follows:

Correspondence contributions are short comments on topical issues, anecdotal material or reactions of readers to nontechnical material published in *Nature* (mainly News, Briefings, Opinion, Book Review or Commentary articles). All are unsolicited; submissions by e-mail to corres@nature.com are preferred (accents on names and places should be clearly indicated).

Commentary articles are informal, editorial comments on topical issues of public interest that have some bearing on scientific research. They are often controversial. Unsolicited manuscripts are considered, but enquiries can be made to the Commentary Editor by e-mail (nature@nature.com) or fax as a one-page synopsis or informal letter before formal submission.

News and Views articles inform nonspecialist readers about new scientific advances, either as reported in published papers or in the form of meeting reports. Most are commissioned, but proposals can be made to the News and Views Editor in advance of publication of the paper or of the meeting concerned. Authors are not allowed to discuss their own work or work from elsewhere in their institution. Detailed guidelines are available on request.

Scientific Correspondence is a predominantly peer-reviewed section of the journal specializing in very brief reports of general interest and discussion of topical issues. Priority is given to contributions of fewer than 500 words. Detailed guidelines are available on request. All contributions are unsolicited, and should be submitted by mail, not e-mail. Copies of contributions submitted by fax should be sent by mail with original figures and a cover letter noting that the manuscript has been faxed.

Book Reviews are all commissioned by the Book Review Editor. Unsolicited contributions are not considered.

Nature's editors will request final copy for Commentary, News and Views, Scientific Correspondence and Book Reviews on disk or via e-mail, as appropriate.

A more detailed Guide to Authors is available on *Nature's* WWW sites at <http://www.nature.com> or <http://www.america.nature.com>